

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:25:02 ; Search time 28.15 Seconds
(without alignments)
337.411 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 401
Sequence: 1 MOYDETLIPKRVPSLCSARY.....LIFGEADVQEMAKERKILRL 401

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 188963 seqs, 23686106 residues

Word size : 0

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	100.0	401	1 W69971	Human sodium-depen
2	241	60.1	401	2 W69920	Human haemochromat
3	9	2.2	436	1 W78919	Human haemochromat
4	7	1.7	12	1 R56478	Vitamin E transport
5	7	1.7	20	1 R15569	Immunopeptide #2 d
6	7	1.7	29	1 R31212	HPV-16 E7 peptide.
7	7	1.7	30	1 W93288	Human papillomavir
8	7	1.7	57	1 R66178	Mouse T18 FlC3 il
9	7	1.7	86	1 R72968	Pig kidney cell mu
10	7	1.7	98	1 R22767	HPV E7 peptide. Im
11	7	1.7	98	1 R42361	Human papillomavir
12	7	1.7	98	1 W46886	Amino acid sequenc
13	7	1.7	172	1 R97562	Human papilloma vi
14	7	1.7	185	1 W93370	Papillomavirus E7/
15	7	1.7	220	1 Y02634	Prot.D1/3-E7-mut(C
16	7	1.7	220	1 Y02631	Protdhrl26-E7-His
17	7	1.7	239	1 Y02636	CLYR-E7-His prote
18	7	1.7	253	1 W81586	CTLA4-E7 fusion pr
19	7	1.7	253	1 W87562	Amino terminal CTL
20	7	1.7	253	1 Y01502	Amino terminal CTL
21	7	1.7	253	1 W97612	HPV 16 E7 protein
22	7	1.7	262	1 R27724	Human papilloma vi
23	7	1.7	266	1 R97561	Porcine papilloma vi
24	7	1.7	341	1 R70142	Pig kidney cell mu
25	7	1.7	341	1 W26964	Human chemokine re
26	7	1.7	344	1 W26767	Amino acid sequenc
27	7	1.7	344	1 W23957	Human macrophage/d
28	7	1.7	356	1 W48087	Prot.D1/3-E6-E7-H1
29	7	1.7	371	1 Y02633	CLYR-E6E7-His pro
30	7	1.7	390	1 W02637	Human brain sodium
31	7	1.7	560	1 W05148	Human sodium-11thl
32	7	1.7	560	1 W70500	Human sodium-11thl
33	7	1.7	576	1 W88523	Eat-4 protein amin

34	7	1.7	671	1 R85290	Streptococcus faec
35	6	1.5	11	1 R33150	HPV E7 protein - R
36	6	1.5	13	1 R10628	Human Papilloma Vi
37	6	1.5	20	1 W40251	Human wild-type E-
38	6	1.5	20	1 W36605	Human E-Cadherin v
39	6	1.5	22	1 W06566	Human preprocollag
40	6	1.5	24	1 R33151	HPV E7 protein - R
41	6	1.5	33	1 W40008	Peptide effecting
42	6	1.5	40	1 R96006	Asymmetry sequence
43	6	1.5	47	1 P94146	Amino acid sequenc
44	6	1.5	64	1 Y07882	Human secreted pro
45	6	1.5	67	1 W73466	Human secreted pro

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description
1	W69971	401	100.0	401	1 W69971	Human sodium-depen
2	W69920	241	60.1	401	2 W69920	Human haemochromat
3	W78919	9	2.2	436	1 W78919	Human haemochromat
4	R56478	7	1.7	12	1 R56478	Vitamin E transport
5	R15569	7	1.7	20	1 R15569	Immunopeptide #2 d
6	R31212	7	1.7	29	1 R31212	HPV-16 E7 peptide.
7	W93288	7	1.7	30	1 W93288	Human papillomavir
8	R66178	7	1.7	57	1 R66178	Mouse T18 FlC3 il
9	R72968	7	1.7	86	1 R72968	Pig kidney cell mu
10	R22767	7	1.7	98	1 R22767	HPV E7 peptide. Im
11	R42361	7	1.7	98	1 R42361	Human papillomavir
12	W46886	7	1.7	98	1 W46886	Amino acid sequenc
13	R97562	7	1.7	172	1 R97562	Human papilloma vi
14	W93370	7	1.7	185	1 W93370	Papillomavirus E7/
15	Y02634	7	1.7	220	1 Y02634	Prot.D1/3-E7-mut(C
16	Y02631	7	1.7	220	1 Y02631	Protdhrl26-E7-His
17	Y02636	7	1.7	239	1 Y02636	CLYR-E7-His prote
18	W81586	7	1.7	253	1 W81586	CTLA4-E7 fusion pr
19	W87562	7	1.7	253	1 W87562	Amino terminal CTL
20	Y01502	7	1.7	253	1 Y01502	Amino terminal CTL
21	W97612	7	1.7	253	1 W97612	HPV 16 E7 protein
22	R27724	7	1.7	262	1 R27724	Human papilloma vi
23	R97561	7	1.7	266	1 R97561	Porcine papilloma vi
24	R70142	7	1.7	341	1 R70142	Pig kidney cell mu
25	W26964	7	1.7	341	1 W26964	Human chemokine re
26	W26767	7	1.7	344	1 W26767	Amino acid sequenc
27	W23957	7	1.7	344	1 W23957	Human macrophage/d
28	W48087	7	1.7	356	1 W48087	Prot.D1/3-E6-E7-H1
29	Y02633	7	1.7	371	1 Y02633	CLYR-E6E7-His pro
30	W02637	7	1.7	390	1 W02637	Human brain sodium
31	W05148	7	1.7	560	1 W05148	Human sodium-11thl
32	W70500	7	1.7	560	1 W70500	Human sodium-11thl
33	W88523	7	1.7	576	1 W88523	Eat-4 protein amin

Db 121 ILGGFISFLGMPFVEYFYGWGCVCCLLMFVYIDDPVSPYPMWISTEKEYIISLKQ 180
 QY 181 VGSSKQPLPIKAMRLSPIMISICGCSHOMVSTWVYIPTYSYHNIRDNGLLSA 240
 Db 181 VGSSKQPLPIKAMRLSPIMISICGCSHOMVSTWVYIPTYSYHNIRDNGLLSA 240
 QY 241 LPTIVAMVIGWGGYIADFLTKRFRITVRKTIATIGSPSSALIVSLPYLNSGYTAT 300
 Db 241 LPTIVAMVIGWGGYIADFLTKRFRITVRKTIATIGSPSSALIVSLPYLNSGYTAT 300
 QY 301 ALLTSCGSLTSCGGYIYINVDIAPRYSFSLMGASRGSSIAPIVYVSGFLSQDPE 360
 Db 301 ALLTSCGSLTSCGGYIYINVDIAPRYSFSLMGASRGSSIAPIVYVSGFLSQDPE 360
 QY 361 FGMRNVFFLLFAVNLGLFYLIFGEADVQEMAKERRKRL 401
 Db 361 FGMRNVFFLLFAVNLGLFYLIFGEADVQEMAKERRKRL 401

RESULT 2
 W78920
 ID W78920 standard; Protein; 401 AA.
 AC W78920;
 DT 21-DEC-1998 (first entry)
 DE Human haemochromatosis protein NTP4.
 KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
 KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 KW type 1 sodium transport gene.
 OS Homo sapiens.
 PN WO9814466-A1.
 PD 09-APR-1998.
 PF 30-SEP-1997; U17658.
 PR 07-MAY-1997; US-852495.
 PR 01-OCT-1996; US-724394.
 PA (PROG-) PROGENITOR INC.
 PI Feder JN, Kronmal GS, Laufer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 DR N-PSDB: V57910.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Claim 60: Fig 5B; 209pp; English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the HFE gene mutation indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophilin (BTF), and can be used in the production of agonists
 CC and antagonists of BTF function. Also described are: (1) a Roret gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence represents
 CC NTP4.
 SO Sequence 401 AA;

Query Match 60.1%; Score 241; DB 1; Length 401;
 Best Local Similarity 100.0%; Pred. No. 9.7e-237;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 SYPMISTSEKEYIISLKQVSSKQPLPIKAMRLSPIMISICGCSHOMVSTWVYI 220
 Db 161 SYPMISTSEKEYIISLKQVSSKQPLPIKAMRLSPIMISICGCSHOMVSTWVYI 220

QY 221 PVISSYVHNIRDNGLLSALPEIVAMVIGWGGYIADFLTKRFRITVRKTIATIGSL 280
 Db 221 PVISSYVHNIRDNGLLSALPEIVAMVIGWGGYIADFLTKRFRITVRKTIATIGSL 280
 QY 281 PSSALLIVSLPYLNSGYTATATLITSCGSLTSCGGYIYINVDIAPRYSFSLMGASRGFS 340
 Db 281 PSSALLIVSLPYLNSGYTATATLITSCGSLTSCGGYIYINVDIAPRYSFSLMGASRGFS 340
 QY 341 SIAPVIVPVSGFLSQDPEFGMRNVFFLLFAVNLGLFYLIFGEADVQEMAKERRKRL 400
 Db 341 SIAPVIVPVSGFLSQDPEFGMRNVFFLLFAVNLGLFYLIFGEADVQEMAKERRKRL 400
 QY 401 L 401
 Db 401 L 401

RESULT 3
 W78919
 ID W78919 standard; Protein; 436 AA.
 AC W78919;
 DT 21-DEC-1998 (first entry)
 DE Human haemochromatosis protein NTP3.
 KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
 KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 KW type 1 sodium transport gene.
 OS Homo sapiens.
 PN WO9814466-A1.
 PD 09-APR-1998.
 PF 30-SEP-1997; U17658.
 PR 07-MAY-1997; US-852495.
 PR 01-OCT-1996; US-724394.
 PA (PROG-) PROGENITOR INC.
 PI Feder JN, Kronmal GS, Laufer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 DR N-PSDB: V57909.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Claim 54: Fig 5B; 209pp; English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the HFE gene mutation indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophilin (BTF), and can be used in the production of agonists
 CC and antagonists of BTF function. Also described are: (1) a Roret gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence represents
 CC NTP3.
 SO Sequence 436 AA;

Query Match 2.2%; Score 9; DB 1; Length 436;
 Best Local Similarity 100.0%; Pred. No. 0.61;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 GCVCCLLMF 152
 Db 220 GCVCCLLMF 228

RESULT 4

R56478 R56478 standard; peptide; 12 AA.
ID R56478:
AC R56478:
DE 12-SEP-1994 (first entry)
DE Vitamin E transport protein fragment.
KW Vitamin E; transport protein.
OS Rattus rattus.
PN J06025299-A.
PD 01-FEB-1994.
PR 23-MAY-1991; JP-146564.
PA (EISA) EISAI CO. LTD.
DR WPI; 94-071958/09.
PT Vitamin-E-specific transport protein(s) - are effective
PT transporters of vitamin-E
PS Claim 1; Page 8; 8pp; Japanese.
CC This is the partial sequence of a vitamin E transport protein. The
CC transport protein also comprises the partial sequence described in
CC R48297.
SQ Sequence 12 AA;

Query Match 1.7%; Score 7; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 NSTSPOS 55
DB 3 NSTSPOS 9

RESULT 5
ID R15569 R15569 standard; protein; 20 AA.
AC R15569:
DE 02-MAR-1992 (first entry)
DE Immunopeptide #2 derived from HPV16 E7 peptide.
KW cervical cancer; cervical intraepithelial neoplasia; CIN; wart;
KW squamous cell carcinoma; ELISA; HPV 16.
OS Synthetic.
PN W09118294-A.
PD 28-NOV-1991.
PR 13-MAY-1991; SE0335.
PA (MEDS-) MEDSCAND AB.
PI Dillner J, Dillner L, Cheng HM;
DR WPI; 91-369390/50.
PT Diagnosis of human papilloma virus infection and PV-carrying
PT tumours - using synthetic peptide(s) to detect virus specific
PT antigen-antibody complexes by immunoassay
PS Disclosure; Page 38; 72pp; English.
CC This is one of two peptides which have been synthesised on the
CC basis of the amino acid sequence for the E7 protein of HPV 16. The
CC selection of peptide sequences was based on the assumption that an
CC immunoreactive region might be situated in the same relative region
CC of a protein from different HPV types. The peptides were used in
CC diagnostic immunoassays to detect HPV-infection.
CC See R15523-R15601.
SQ Sequence 20 AA;

Query Match 1.7%; Score 7; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
DB 11 QLNDSSE 17

RESULT 6
ID R31212 R31212 standard; peptide; 29 AA.

AC R31212:
DE 19-MAY-1993 (first entry)
DE HPV-16 E7 peptide.
DE Human papillomavirus; E7; epitope; cervical cancer; invasive;
KW antigens; diagnostic; cytotoxic; tagged.
OS Human papillomavirus type 16.
PN EP-523391-A.
PD 20-JAN-1993.
PR 19-JUN-1992; 110367.
PR 13-JUL-1991; EP-111720.
PA (BEHW) BEHRINGWERKE AG.
PI Gissmann L, Mueller M;
DR WPI; 93-019474/03.
PT Use of HPV-16 E6 and E7-gene derived peptide(s) and their specific
PT antibodies - for treatment and diagnosis of HPV-16
PT associated invasive cervical cancer
PS Disclosure; Page 7; 14pp; English.
CC The peptide represents an epitope of the human papillomavirus type 16
CC E7 protein, from amino acids 6-35. This epitope is useful as a
CC target for diagnosis and imaging of HPV-associated cancers, such as
CC invasive cervical cancer. Antibodies raised against this epitope,
CC tagged with cytotoxic molecules, such as cholera toxin, have
CC therapeutic potential.
CC See also R31213-5.
SQ Sequence 29 AA;

Query Match 1.7%; Score 7; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
DB 21 QLNDSSE 27

RESULT 7
ID W93288 W93288 standard; peptide; 30 AA.
AC W93288:
DE 27-MAY-1999 (first entry)
DE Human papillomavirus peptide fragment #2.
KW Diagnosis; skin; immune reaction; onco-protein; E6; E7.
OS Human papillomavirus.
PN DE19737409-A1.
PD 04-MAR-1999.
PR 27-AUG-1997; 037409.
PR 27-AUG-1997; DE-037409.
PA (MEDT-) MEDIGENE AG.
PI Hoepfl R;
DR WPI; 99-168276/15.
PT Diagnosis kit for testing skin for immune reactions against
PT onco-protein E6 and E7 - comprises onco-protein E6 and E7 and/or
PT immunologically active parts of E6 and E7 derived from human
PT papilloma virus
PS Disclosure; Column 3; 4pp; German.
CC This invention describes peptides used in a diagnosis kit for testing
CC skin for immune reactions against onco-protein E6 and E7. The method
CC of the invention comprises onco-protein E6 and E7 and/or immunologically
CC active parts of E6 and E7 derived from human papilloma virus.
SQ Sequence 30 AA;

Query Match 1.7%; Score 7; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
DB 7 QLNDSSE 13

RESULT 8

R6178
ID R6178 standard; Peptide; 57 AA.
AC R6178;
DT 10-AUG-1995 (first entry)
DE Mouse T118 Flt3 ligand peptide fragment.
KW Flt3 ligand; tyrosine kinase receptor ligand; ss.
OS Mus musculus.
PN W09426891-A.
PD 24-NOV-1994.
PE 18-MAY-1994; U05150.
PR 19-MAY-1993; US-065231.
PR 07-JUL-1993; US-089263.
PR 16-JUL-1993; US-092543.
PR 13-AUG-1993; US-106340.
PR 24-AUG-1993; US-112391.
PR 19-NOV-1993; US-155111.
PR 03-DEC-1993; US-162413.
PA (INRM) INSR NAT SANTE & RECH MEDICALE.
PI (SCHE) SCHERING CORP.
PI Birmahum D, Calpepper JA, Hannum CH, Lee FD;
DR WPI; 95-006787/01.
DR N-PSDB; Q79466.
PT New ligand for the Flt3 tyrosine kinase receptor - and related
PT nucleic acid, vectors, host cells and antibodies, useful for
PT treating abnormal cell physiology and proliferation, e.g. cancer,
PT also for diagnosis and drug screening
PS Claim 11; page 80; 90pp; English.
CC A cDNA library from the human stromal cell line 29SV48, in
CC pME18S, was screened with an 800 bp fragment derived from
CC mouse clone T118. This fragment encompasses the coding region
CC conserved between two mouse clones, T118 and T110. Approx. 20
CC positive clones were selected and partially sequenced. Two
CC clones, S86 and S109, were found to be approx. 75% homologous
CC to the mouse clones over the first 163 AAs. Clone S86 continued
CC to show homology to T110 until the stop codon, although to a
CC lesser degree, for an overall homology of 66%. Clones T118 and
CC S109 do not show homology to each other or to the other clones
CC after mouse residue 163 (human residue 160). An additional mouse
CC clone designated MB8 has a 29 AA insert at the junction between
CC the common and divergent portions of the mouse ligand.
SQ Sequence 57 AA;

Query Match 1.7%; Score 7; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 TATALLT 304
DB 16 TATALLT 22

RESULT 9
ID R72968 standard; Peptide; 86 AA.
AC R72968;
DT 10-JAN-1996 (first entry)
DE Pig kidney cell mutarotase peptide fragment IX-15.
KW Mutarotase; pig; kidney; aldose-1 epimerase; catalytic conversion;
KW glucose; lambda library; probe; pBluescript; E.coli.
PN Sus scrofa.
PD J06253856-A.
PD 13-SEP-1994.
PE 26-FEB-1993; 062646.
PR 26-FEB-1993; JP-062646.
PA (AMAN) AMANO PHARM KK.
DR WPI; 95-181575/24.
PT Mutarotase gene - useful for the efficient mass production of
PT mutarotase
PS Example 1; Page 6; 7pp; Japanese.
CC The amino acid sequences (R72965-8) are internal protease digested
CC fragments of the mutarotase protein from pig kidney cells (R72966).
CC Mutarotase is an aldose-1 epimerase which catalyses the conversion

CC of alpha-D-glucose to beta-D-glucose. This fragment corresponds to
CC amino acids 145-231 of the entire protein. The peptides were generated
CC by digestion with trypsin, chymotrypsin, endopeptidase and Staphylococcus
CC V8 protease. The gene was isolated from a pig kidney cell cDNA lambda
CC library using the probe Q85558. The gene was inserted into the plasmid
CC pBluescript II and transformed into E.coli JM109 for production of the
CC protein.
SQ Sequence 86 AA;

Query Match 1.7%; Score 7; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VDETLIP 9
DB 59 VDETLIP 65

RESULT 10
ID R22767 standard; peptide; 98 AA.
AC R22767;
DT 21-SEP-1992 (first entry)
DE HPV E7 peptide.
KW Human; papillomavirus; immunogenic; cervical; warts; carcinoma;
KW cancer.
OS Synthetic.
OS Homo sapiens.
PN W09205248-A.
PD 02-APR-1992.
PE 26-SEP-1991; U07081.
PR 26-SEP-1990; US-588384.
PA (BRIM) BRISTOL-MYERS SQUIB.
PI Blake J, Chen L, Hellstrom I, Hellstrom K, Hu S L,
PI Thomas E K;
DR WPI; 92-132119/16.
PT Immunogenic peptide(s) derived from E6 or E7 region of HPV16 -
PT and recombinant cells encoding them, useful in treatment and
PT prophylaxis of cervical warts or cancer resulting from HPV
PT infection
PS Disclosure; Fig 7; 81pp; English.
CC The peptide is the sequence of the human papillomavirus HPV 16 E7
CC nucleoprotein. Peptides corresponding to regions (pref. epitopic
CC regions) of HPV 16 E7 were synthesised by standard Merrifield
CC synthesis. Examples of such peptides are E7 1-10, 29-50 or 70-81.
CC Compositions contg. these peptides, antibodies against the peptides,
CC or recombinant cells contg. the gene encoding the immunogenic
CC peptides may be utilised in methods for inhibiting and treating HPV
CC infection and tumour initiation and progression e.g. in the
CC prevention or retardation of cervical warts and cervical carcinoma
CC resulting from HPV infection.
CC See also R22766.
SQ Sequence 98 AA;

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
DB 27 QLNDSSE 33

RESULT 11
ID R42361 standard; protein; 98 AA.
AC R42361;
DT 21-MAY-1994 (first entry)
DE Human papillomavirus 16 E7 protein and fragments.
KW Tumours; cows; horses; donkeys; regression; udder warts; HPV16.
OS Synthetic.

PN W09320844-A.
PD 28-OCT-1993.
PF 01-APR-1993; G00679.
PR 08-APR-1992; GB-007701.
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI Campo MS;
DR WPI: 93-351368/44.
PT Use of papilloma-virus E7 protein or fragments for the therapy of
PT papilloma-virus disease - for the regression of tumours e.g.
PT removal of warts from udders or mouth of milking cows or for
PT treatment of horses or donkeys
PS Disclosure; Fig 2; 31pp; English.
CC The sequence is that of the human papillomavirus type 16 E7 protein.
CC The protein sequence was aligned with that of bovine papillomavirus
CC type 4.
CC See also R42360.
SQ Sequence 98 AA;

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
|||||||
DB 27 QLNDSSE 33

RESULT 12
W46886
ID W46886 standard; Protein; 98 AA.
AC W46886;
DT 15-JUN-1998 (first entry)
DE Amino acid sequence of the HPV-16 E7 oncoprotein.
KW E7 oncoprotein; proliferative state; HPV; kinase activity;
KW cyclin/cyclin-dependent kinase; p21CIP1; interaction; inactivation;
OS Human papillomavirus.
PN US5736318-A.
PD 07-APR-1998.
PE 17-MAR-1995; 406248.
PR 17-MAR-1995; US-406248.
PA (HARD) HARVARD COLLEGE.
PI (HARD) UNITV HARVARD.
PI Jones DL, Munger K;
DR WPI: 98-239202/21.
DR N-PSDB: V16717.
PT Evaluation of proliferative state of cells transformed with human
PT papilloma virus - by determining cyclin-dependent kinase activity
PT induced by E7 onco-protein
PS Disclosure; Columns 19-20; 14pp; English.
CC The present sequence represents Human papillomavirus (HPV), strain 16,
CC E7 oncoprotein. The proliferative state of a cell transformed with
CC HPV can be evaluated in the following manner. Cyclin/cyclin-dependent
CC kinase complexes containing protein p21CIP1 (W46887-88) are isolated
CC from the transformed cell, and the HPV E7 oncoprotein added to the
CC isolated protein. Cyclin/cyclin-dependent kinase complexes are isolated
CC from an untransformed cell that is substantially homogenic with the
CC transformed cell, and the HPV E7 oncoprotein added. The kinase activities
CC of the 2 samples are measured, where a proliferating transformed cell
CC has a greater kinase activity than the untransformed cell. The method is
CC used for determining the extent of interaction and/or inactivation
CC between a cyclin/cyclin-dependent kinase inhibitor and the HPV E7
CC oncoprotein and thus evaluating the proliferative state of a transformed
CC cell.
SQ Sequence 98 AA;

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62

DB 27 QLNDSSE 33
|||||||

RESULT 13
R97562
ID R97562 standard; Protein; 172 AA.
AC R97562;
DT 11-JAN-1997 (first entry)
DE Human papilloma virus E6/E7 protein variant.
KW Human papilloma virus; E6; E7; deletion mutant; HPV;
KW immune response; humoral immune response; cellular immune response;
KW vaccine.
OS Human papilloma virus.
PN W09619496-A1.
PD 27-JUN-1996.
PE 20-DEC-1995; A00868.
PR 20-DEC-1994; AU-000157.
PA (CSLC-) CSL LTD.
PI (UYOU) UNITV QUEENSLAND.
PI Cox J, Edwards SJ, Frazer I, Webb EA;
DR WPI: 96-309518/31.
DR N-PSDB: T31834.
PT Vaccine variants of human papilloma virus antigens - contain
PT variants of E6 and/or E7 protein, pref. deletion mutants, and are
PT used to treat or prevent HPV infection
PS Example 3; Page 17; 37pp; English.
CC A variant of the human papilloma virus (HPV) E6 or E7 protein which
CC elicits a humoral and/or cellular immune response against HPV can be
CC used in vaccines against HPV or to treat HPV infection. The variant
CC is preferably a deletion mutant comprising at least half, and
CC preferably two-thirds of full length E6 or E7 protein starting from
CC the N- or C-terminal, or is a full length E6 moiety fused to a full
CC length E7 moiety. The variant optionally has a linkage moiety and a
CC foreign protein or peptide which facilitates the purification of,
CC and enhances the immunogenicity of, the fusion protein. This
CC is a fusion protein of the C-terminal end of E6 and the
CC N-terminal end of E7. The protein is also a deletion mutant
CC generated from the sequence described in T31833.
SQ Sequence 172 AA;

Query Match 1.7%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
|||||||
DB 125 QLNDSSE 131

RESULT 14
W99370
ID W99370 standard; peptide; 185 AA.
AC W99370;
DT 21-MAY-1999 (first entry)
DE Papillomavirus E7/glycoprotein signal peptide fusion protein.
KW Antitumour; immunogen; intracellular localisation; cancer; dysplasia;
KW signal peptide; papillomavirus; infection; cervix uteri.
OS Chimeric - Human papillomavirus.
OS Chimeric - Rabies virus.
PN W09903885-A1.
PD 28-JAN-1999.
PE 17-JUL-1998; F01576.
PR 18-JUL-1997; FR-009152.
PA (TRGE) TRANSGENE SA.
PI Balloul JM, Bizouarne N, Kieny MP;
DR WPI: 99-132161/11.
PT Antitumour composition containing immunogenic polypeptide with
PT altered localisation - or vector expressing this polypeptide,
PT particularly for treating or preventing cervical cancer associated
PT with human papilloma virus
PS Claim 16; Page 55-56; 74pp; French.

CC The invention relates to an antitumour composition containing a
 CC therapeutic or prophylactic agent or one or more immunogenic
 CC polypeptides, where at least one polypeptide is modified so that
 CC its intracellular localisation is different from that of the native
 CC polypeptide. This sequence represents an example of a modified
 CC immunogenic protein and corresponds to the E7 protein from human
 CC papillomavirus fused to the rabies virus glycoprotein signal peptide.
 CC The compositions, vectors and particles are used to treat or prevent
 CC cancers and tumours, specifically those associated with papillomavirus
 CC infection (e.g. cancer or low grade dysplasia of the cervix uteri).
 SQ Sequence 185 AA;

Query Match 1.7%; Score 7; DB 1; Length 185;
 Best Local Similarity 100.0%; Pred. No. 29;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
 |||||

DB 46 QLNDSSE 52

RESULT 15

Y02634
 ID Y02634 standard; Protein; 220 AA.
 AC Y02634;
 DT 22-JUN-1999 (first entry)
 DE Prot.D1/3-E7-mult(C24G,E26Q)/HPV16 protein.
 KW Chimeric; E6; E7; fusion protein; protein D; vaccine; immunotherapy;
 KM tumour; lesion; benign; malignant; virus; infection.
 OS Chimeric - Human Papillomavirus.
 PN WO910375-A2.
 PD 04-MAR-1999.
 PF 17-AUG-1998; E05285.
 PR 22-AUG-1997; GB-017953.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Bruck C, Cabezon Silva T, Delisse AER, Gerard CMG,
 PI Lombardo-Bencheikh A;
 DR WPI: 99-190587/16.
 DR N-PSDB; X29783.
 DR Human Papilloma Virus (HPV) fusion proteins - useful in vaccines for
 PT treatment or prophylaxis of HPV induced lesions
 PS Disclosure: Fig 8; 95pp; English.
 CC This sequence represents a chimeric E6 or E7 protein or E6/E7 fusion
 CC protein from Human Papillomavirus (HPV) linked to an immunological fusion
 CC partner, in this case, a fragment of the Haemophilus influenzae B
 CC protein D. The sequence also contains a histidine tag at the C-terminus
 CC of the encoded protein. The protein can be used in a vaccine, for
 CC immuno-therapeutically treating HPV induced tumour lesions (benign or
 CC malignant) and preventing HPV viral infection.
 SQ Sequence 220 AA;

Query Match 1.7%; Score 7; DB 1; Length 220;
 Best Local Similarity 100.0%; Pred. No. 34;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
 |||||

DB 140 QLNDSSE 146

Search completed: May 25, 2000, 14:54:02
 Job time: 1740 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OW protein - protein search, using sw model

Run on: May 25, 2000, 14:27:18 ; Search time 40.28 Seconds

(without alignments)
143.729 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 401

Sequence: 1 MOVDELIPKRVPSICSAARY.....LIFGEADVQEWAKRKRLRL 401

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 145308 seqs, 14437401 residues

Word size : 0

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/6.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6.COMB.pep:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	100.0	401	2	US-08-805-118-1
2	138	34.4	480	2	US-08-724-394A-11
3	9	2.2	470	2	US-08-724-394A-10
4	7	1.7	20	2	US-08-934-915-47
5	7	1.7	30	1	US-08-363-586-1
6	7	1.7	57	4	PCT-US94-05150-22
7	7	1.7	98	1	US-08-406-248-6
8	7	1.7	253	2	US-08-459-818-20
9	7	1.7	253	2	US-08-889-666-20
10	7	1.7	253	2	US-08-465-078-20
11	7	1.7	253	2	US-08-725-776-20
12	7	1.7	253	2	US-08-488-062-20
13	7	1.7	263	1	US-08-117-083-9
14	7	1.7	467	1	US-08-805-118-3
15	7	1.7	480	2	US-08-724-394A-9
16	7	1.7	560	1	US-08-647-484-2
17	7	1.7	560	1	US-08-647-481-2
18	7	1.7	560	1	US-08-430-033A-2
19	7	1.7	560	1	US-08-805-118-4
20	7	1.7	560	4	PCT-US96-05792-2
21	7	1.7	671	1	US-08-737-716-13
22	6	1.5	22	1	US-08-466-265-2
23	6	1.5	22	1	US-08-755-728-15
24	6	1.5	22	2	US-08-974-655-15
25	6	1.5	40	2	US-08-353-476-92
26	6	1.5	100	2	US-08-465-380-20
27	6	1.5	100	2	US-08-480-478-48
28	6	1.5	100	2	US-08-486-397-20

29	6	1.5	100	2	US-08-486-399-20	Sequence 20, Appl
30	6	1.5	100	2	US-08-461-965-20	Sequence 20, Appl
31	6	1.5	100	2	US-08-326-110A-48	Sequence 48, Appl
32	6	1.5	100	2	US-08-634-641-20	Sequence 20, Appl
33	6	1.5	100	3	US-09-249-471-20	Sequence 20, Appl
34	6	1.5	116	1	US-08-053-131-185	Sequence 185, App
35	6	1.5	116	2	US-08-096-762-185	Sequence 185, App
36	6	1.5	118	4	PCT-US93-07213-13	Sequence 113, Appl
37	6	1.5	160	1	US-08-479-223-11	Sequence 11, Appl
38	6	1.5	160	4	PCT-US93-00643-11	Sequence 11, Appl
39	6	1.5	190	1	US-08-441-629-16	Sequence 16, Appl
40	6	1.5	190	4	PCT-US95-09172-16	Sequence 16, Appl
41	6	1.5	193	2	US-08-438-439C-21	Sequence 21, Appl
42	6	1.5	195	1	US-08-236-427-4	Sequence 4, Appl
43	6	1.5	200	4	PCT-US93-07213-7	Sequence 7, Appl
44	6	1.5	205	1	US-08-340-820-5	Sequence 5, Appl
45	6	1.5	205	1	US-08-172-328-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-805-118-1
; Sequence 1, Application US/08805118
; Patent No. 5985604
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
; TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,118
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ballings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0221 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITV02
; CLONE: 754412
; US-08-805-118-1

Query Match 100.0%; Score 401; DB 2; Length 401;
Best local Similarity 100.0%; Pred. No. 0;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOVDETLIPKRVPSLCASRGIALVLFHCNFTTIAQNVIMITWAVANSTPOSOLNDS 60
Db 1 MOVDETLIPKRVPSLCASRGIALVLFHCNFTTIAQNVIMITWAVANSTPOSOLNDS 60
QY 61 SEVLPVDSFGSLKAPRSPLPAKSSILGQFAIMERMGPPOERSHLCIALSGMLGCF7A 120
Db 61 SEVLPVDSFGSLKAPRSPLPAKSSILGQFAIMERMGPPOERSHLCIALSGMLGCF7A 120
QY 121 ILIGFSETLGMPEVFIYIGVGVCCCLMFVYIDDPVSYPMISSEKEYITISLKQ 180
Db 121 ILIGFSETLGMPEVFIYIGVGVCCCLMFVYIDDPVSYPMISSEKEYITISLKQ 180
QY 181 VGSSKQPLPKAMLRSLPMSICLGCFSHOMLVSTWVVIPTIISYVHNINDNGLLSA 240
Db 181 VGSSKQPLPKAMLRSLPMSICLGCFSHOMLVSTWVVIPTIISYVHNINDNGLLSA 240
QY 241 LPTIVAVIOMVGGYLADELTKKRLITRKATITIGSLPSALLVSLPYLNSGYTTAT 300
Db 241 LPTIVAVIOMVGGYLADELTKKRLITRKATITIGSLPSALLVSLPYLNSGYTTAT 300
QY 301 ALLTSCGLSTLCOGSIYINVDIAPRYSFLMGASRGFSSIAPIVPTVSGFLLSODPE 360
Db 301 ALLTSCGLSTLCOGSIYINVDIAPRYSFLMGASRGFSSIAPIVPTVSGFLLSODPE 360
QY 361 FGRNVFFLLFAVNLGLFYLIFGADVOEWAKERKRLRL 401
Db 361 FGRNVFFLLFAVNLGLFYLIFGADVOEWAKERKRLRL 401

RESULT 2
US-08-724-394A-11
; Sequence 11, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0300
; TELEFAX: 415-576-0200
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..480
; OTHER INFORMATION: /note="NP74"
US-08-724-394A-11

Query Match 34.4%; Score 138; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 5.5e-123;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 KRLLIVRKIAITIGSLPSALLVSLPYLNSGYTTATALLTSCGLSTLCOGSIYINVD 323
Db 343 KRLLIVRKIAITIGSLPSALLVSLPYLNSGYTTATALLTSCGLSTLCOGSIYINVD 402
QY 324 IAPRYSFLMGASRGFSSIAPIVPTVSGFLLSODPEGRWVFFLLFAVNLGLFYL 383
Db 403 IAPRYSFLMGASRGFSSIAPIVPTVSGFLLSODPEGRWVFFLLFAVNLGLFYL 462
QY 384 FGEADVOEWAKERKRLRL 401
Db 463 FGEADVOEWAKERKRLRL 480

RESULT 3
US-08-724-394A-10
; Sequence 10, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0300
; TELEFAX: 415-576-0200
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region

Mon Jun 12 23:02:41 2000

LOCATION: 1.470
OTHER INFORMATION: /note= "NPT3"
US-08-724-394A-10

Query Match 2.2%; Score 9; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 GCVCCLMF 152
Db 220 GCVCCLMF 228

RESULT 4
US-08-934-915-47
Sequence 47, Application US/08934915
Patent No. 5932412
GENERAL INFORMATION:
APPLICANT: DILLNER, JOAKIM
APPLICANT: DILLNER, LENA
APPLICANT: CHENG, HWEE-MING
TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
TITLE OF INVENTION: PAPILOMAVIRUS 1, 5, 6, 8,
TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
TITLE OF INVENTION: USEFUL IN IMMUNOSSAY FOR
TITLE OF INVENTION: DIAGNOSTIC PURPOSES
NUMBER OF SEQUENCES: 193
CORRESPONDENCE ADDRESS:
ADDRESSEE: MASON & ASSOCIATES, P.A.
STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
CITY: CLEARWATER
STATE: FLORIDA
COUNTRY: U.S.A.
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 3.0
SOFTWARE: Microsoft Word 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934, 915
FILING DATE: 22-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/949, 836
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: LOUISE A. Foutch
REGISTRATION NUMBER: 37,133
REFERENCE/DOCKET NUMBER: 1946.6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 813-538-3800
TELEFAX: 813-538-3820
TELEX:
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-934-915-47

Query Match 1.7%; Score 7; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSE 62
Db 11 QLNDSE 17

RESULT 5

US-08-363-586-1
Sequence 1, Application US/08363586
Patent No. 5629161
GENERAL INFORMATION:

APPLICANT: Mueller, Martin
APPLICANT: Glassmann, Lutz
TITLE OF INVENTION: Use of HPV-16 E6 and E7-Gene Derived
TITLE OF INVENTION: Peptides for the Diagnostic Purpose
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,586
FILING DATE: 23-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/909,296
FILING DATE: 09-JUL-1992
APPLICATION NUMBER: EP 91111720.8
FILING DATE: 13-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hadler, Linda A.
REGISTRATION NUMBER: 33,218
REFERENCE/DOCKET NUMBER: 02481-1195-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-363-586-1

Query Match 1.7%; Score 7; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSE 62
Db 22 QLNDSE 28

RESULT 6
PCT-US94-05150-22
Sequence 22, Application PC/TUS9405150
GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: Purified Mammalian Flt3 Ligands and Agonists and Antagonist
NUMBER OF SEQUENCES: 37
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/05150
FILING DATE:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/162,413
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,111
FILING DATE: 19-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/112,391
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/106,340
FILING DATE: 13-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/092,549
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,263
FILING DATE: 07-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/065,231
FILING DATE: 19-MAY-1993
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 57 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US94-05150-22

Query Match 1.7%; Score 7; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 TATALLT 304
Db 16 TATALLT 22

RESULT 7
US-08-406-248-6
Sequence 6, Application US/08406248
Patent No. 5736318
GENERAL INFORMATION:
APPLICANT: Munger, Karl
APPLICANT: Jones, D. Leanne
TITLE OF INVENTION: METHOD AND KIT FOR EVALUATING
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ann-Louise Kerner, Ph.D., Lappin & Kusner
STREET: 200 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/406,248
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: McDaniel, Patricia A.
REGISTRATION NUMBER: 33,194
TELEPHONE: 617-330-1300
TELEFAX: 617-330-1311
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 98 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-406-248-6

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QUNDSE 62
Db 27 QUNDSE 33

RESULT 8
US-08-459-818-20
Sequence 20, Application US/08459818
Patent No. 5851795
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FastSeq 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,818
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436.35US02
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-459-818-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QUNDSE 62
Db 182 QUNDSE 188

RESULT 9
US-08-889-666-20
Sequence 20, Application US/08889666
Patent No. 5885579

GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kienert, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/889,666
FILING DATE: 08-JUL-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-350S01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-9031
TELEFAX: 310-445-1140
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-889-666-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSSE 62
DB 182 QLNDSSE 188

RESULT 10
US-08-465-078-20
Sequence 20, Application US/08465078
Patent No. 5685796
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kienert, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,078
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-350S01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-9031
TELEFAX: 310-445-1140
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-078-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSSE 62
DB 182 QLNDSSE 188

RESULT 11
US-08-725-776-20
Sequence 20, Application US/08725776
Patent No. 5968510
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kienert, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,776
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-350S01

TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-776-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSE 62
|||||
Db 182 QLNDSE 188

RESULT 12
US-08-488-062-20
Sequence 20, Application US/084488062
Patent No. 597318
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Dangle, Milton K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,062
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-35US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-062-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSE 62
|||||
Db 182 QLNDSE 188

RESULT 13
US-08-117-083-9
Sequence 9, Application US/08117083
Patent No. 5719054
GENERAL INFORMATION:
APPLICANT: Bournell, Michael E.
APPLICANT: Inglis, Stephen C.
APPLICANT: Munro, Alan J.
TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
TITLE OF INVENTION: Papilloma Virus Proteins
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Walter H. Dregger
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,083
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dregger, Walter H.
REGISTRATION NUMBER: 24,190
REFERENCE/DOCKET NUMBER: A-58783
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..263
OTHER INFORMATION: /note="Xaa refers to stop codon in
US-08-117-083-9 the open reading frame."

Query Match 1.7%; Score 7; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSE 62
|||||
Db 188 QLNDSE 194

RESULT 14
US-08-805-118-3
Sequence 3, Application US/08805118
Patent No. 5985604
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT

TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,118
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0221 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 450532
US-08-805-118-3

Query Match 1.7%; Score 7; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 LDIAPRY 328
DB 388 LDIAPRY 394

RESULT 15
US-08-724-394A-9
Sequence 9, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laner, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Therefo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 01/957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Region
LOCATION: 1..480
OTHER INFORMATION: /note="NPT1"
US-08-724-394A-9

Query Match 1.7%; Score 7; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 LDIAPRY 328
DB 401 LDIAPRY 407

Search completed: May 25, 2000, 14:54:52
Job time: 1654 sec

Mon Jun 12 23:02:41 2000

us-09-391-958-1.oli.go.ra1

OM of: US-09-391-958-1 to: GenEmbl: * out_format : pfs
Date: May 25, 2000 3:26 PM
About: Results were produced by the GenCore software, version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

```
Search information block:
Query: US-09-391-958-1
Query length: 401
Database: GenEmbl.*
Database sequences: 882769
Database length: -486395729
Search time (sec): 737.360000
```

score_list:

gb_pi4:AC005728	+	9.00	107.43	625.78	131389	!	AC005728	Homo sapiens chr
gb_htg4:AC011093	-	9.00	107.43	626.18	131485	!	AC011093	Homo sapiens chr
gb_htg4:AL138726	-	9.00	106.99	662.52	140366	!	AL138726	Homo sapiens chr
gb_htg12:AC010575	-	9.00	106.60	696.11	148447	!	AC010575	Drosophila melan
gb_htg4:AC012145	-	9.00	106.57	699.00	149155	!	AC012145	Homo sapiens chr
gb_htg1:HS0181N1	+	9.00	106.42	712.87	152550	!	Z82191	Homo sapiens chrom
seq_name: gb_pat:AR036571								
seq_documentation_block:								
LOCUS	AR036571	1780	bp	DNA		PAT	29-SEP-1999	
DEFINITION	Sequence 19 from patent US 5872237.							
ACCESSION	AR036571							
VERSION	AR036571.1	GI:5953239						
KEYWORDS								
SOURCE	Unknown.							
ORGANISM	Unknown.							
REFERENCE	Unclassified.							
AUTHORS	1 (bases 1 to 1780)							
TITLE	Feder, J., Nathan, K., Kroman, G., Scott, Lauer, P. M., Ruddy, D. A., Thomas, W.,							
	Tsuchihashi, Z., and Wolff, R. K.							
	Megabase transcrip map: novel sequences and antibodies thereto							
	Patent: US 5872237-A 19 16-FEB-1999.							
JOURNAL	Location/Qualifiers							
FEATURES	1..1780							
BASE COUNT	458	a	436	c	402	g	484	t
ORIGIN	/organism="unknown"							
Alignment_scores:								
Quality:		241.00	Length:		241			
Ratio:		1.000	Gaps:		0			
Percent Similarity:		100.000	Percent Identity:		100.000			
Alignment_block:								
US-09-391-958-1 x AR036571 ..								
Align seg 1/1 to: AR036571 from: 1 to: 1780								
161	SerTyrProT	P	I	L	S	E	R	Th
177	uLysGlnI	N	A	L	I	G	L	S
857	T	C	C	A	T	T	C	C
177	uLysGlnI	N	A	L	I	G	L	S
907	G	A	A	C	A	A	G	G
194	e	A	u	A	r	G	S	e
957	T	C	A	G	A	T	C	T
211	T	r	p	l	e	u	a	l
1007	T	G	G	T	A	G	C	T
227	I	T	y	r	H	i	s	a
1057	G	T	A	C	A	T	T	A
244	I	e	a	l	a	T	r	p
1107	T	T	G	T	G	C	C	T
261	L	e	u	t	h	r	L	y
1157	C	T	A	A	C	A	A	A
277	u	G	L	y	S	e	r	L
1207	A	G	G	A	G	T	C	C
294	e	r	G	L	y	T	r	L

```

1257 CCGGCTAATACACGCAACTGCCCTGACGCTCTCTGGGATTAAGC 1306
311 ThrLeuCySgInSerGlyIleTyrlLeuValLeuAspIleAlaProAr 327
1307 ACATGTGTCACTCAGTCAGGATTTATATCAATGTCTAGATATGCTCCAG 1356
327 gTySerSerPheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaP 344
1357 GTATTCCAGTTTCTCATGGAGCATCAAGAGATTTTCGAGCATAGCAC 1406
344 roValIleValProThrValSerGlyPheLeuLeuSerGlnAspProGlu 360
1407 CTGTCAATGTACCCACGTCTCAGCGATTTCTTCTAGTCAGACCTGAG 1456
361 PheGlyTrpArgAsnValPhePheLeuLeuPheAlaValAsnLeuLeuG 377
1457 TTTGGGTGAGGAATGCTCTCTCTGCTGTTCCGTTAACCTGTTAGG 1506
377 yLeuLeuPheTyrlLeuIlePheGlyGlyAlaAspValGlnGluTrpAla 394
1507 ACTACTCTTCTACCTCATATTGAGAGACAGATGTCAGAAATGGGCTA 1556
394 ySgIuArgLyLeuThrArgLeu 401
1557 AAGAGAGAAAACCTACCTCGTTTA 1579
seq_name: gb_p13:HSU90545

```

```

seq_documentation_block:
LOCUS HSU90545 1795 bp mRNA PRI 02-MAY-1997
DEFINITION Human sodium phosphate transporter (NP74) mRNA, complete cds.
ACCESSION U90545
VERSION U90545.1 GI:2062691
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

```

REFERENCE

AUTHORS

1 (bases 1 to 1795)
Ruddy,D.A., Krommal,G.S., Lee,V.K., Mintier,G.A., Quintana,L.,
Domingo,R., Jr., Meyer,N.C., Basava,A., McClelland,E., Fullan,A.,
Mapa,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchihashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
A 1.1 megabase transcript map of the human hereditary
hemochromatosis locus
Unpublished

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 1795)
Ruddy,D.A., Krommal,G.S., Lee,V.K., Mintier,G.A., Quintana,L.,
Domingo,R., Jr., Meyer,N.C., Basava,A., McClelland,E., Fullan,A.,
Mapa,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchihashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
Direct Submission
Submitted (25-FEB-1997) Sequencing, Mercator Genetics, 4040
Campbell Avenue, Menlo Park, CA 94025, USA

FEATURES

source

1..1795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
377..1582
/feature="NP74"
377..1582
/feature="NP74"
377..1582
/feature="NP74"
/codon_start=1
/product="sodium phosphate transporter"
/protein_id="AAB53423.1"
/db_xref="GI:2062692"

gene

CDS

translation="MOVDELTPRKPSICSAARYGIALVALHCFNFTTIAQNYINMTM
VAMVNTSPQSLNDSSEVLPAVDSFGLSKAPKSLPAKSSIIAGGFALIEKGGPOER
SRKSTALSGMLGCFETALLIGFISSETIGMFEVYIFGVCVCCCLLMFVYIDDF
STPWISTSEKEYIISIKQVSSKQPIPIKAMLSLPIWISICLGFHOMLVSTIMV

```

YIPYISSVYHVNIRNDGLALPFIYAVVIGMGYLADELTKKRLTYVRKART
IGSLPSALIVSIPYNSGYTATLALTSGLSITQSGIYINVLDIRYSSFLMG
ASRGSSIAPIYVPIVPSGFLISQDEPFQNRNVPFLIRAVNLGLFLYLIREEADVQEM
AKERLTRL"
BASE COUNT 473 a 436 c 402 g 484 t
ORIGIN

```

```

alignment_scores:
    Quality: 241.00      Length: 241
    Ratio: 1.000         Gaps: 0
Percent Similarity: 100.000    Percent Identity: 100.000

```

alignment_block:

US-09-391-958-1 x HSU90545 ..

Align seg 1/1 to: HSU90545 from: 1 to: 1795

```

161 SerTyProThrPheSerThrSerGlyIleTyrlIleSerSerle 177
857 TCCTATTCATGATTAAGACCTCAGAAAAAATACATCATATCTCTT 906
177 uLySgInValGlySerSerIleProLeuProIleLyAlaMetL 194
907 GAACAGACAGGTGGGTCTTTAGCAGCCCTCTCCATCAAGCTATGC 956
194 euArgSerLeuProIleTrpSerIleCysLeuGlyCysPheSerHsgIn 210
957 TCAGATCTTACCCATTTGTCATATGTTTAGGCTTTTCAGCATCA 1006
211 TrpLeuValSerThrMetValTyrlIleProThrTyrlIleSerSerVa 227
1007 TGGTTAGTTAGCAACAATGTTGTATACATACCACTTACATCAGCTCTGT 1056
227 TyrHisValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuProPhe 244
1057 GTACCATGTTATACATCAGACAAATGACCTTATCTGCCCTCTTTA 1106
244 leValAlaTrpValIleGlyMetValGlyTyrlLeuAlaAspPheLeu 260
1107 TTGGTCCCTGGGTCAATGATGTTGAGGAGCTATCTGCAGATTTCTT 1156
261 LeuThrLySylsPheArgLeuIlePheValArgyIleAlaThrIleLe 277
1157 CTACCAAAAAGTTTAGCTCATCTCATCTGAGGAAAATGGCAAAATTT 1206
277 uGlySerLeuProSerSerAlaLeuIleValSerLeuProTyrlLeuAsn 294
1207 AGGAAGTCTCCCTCTTCAGCAGCATTTGTCTCTGCTTACCTCAAT 1256
294 ergTyrlIleThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSer 310
1257 CCGGCTATATACAGCAACTGCTTGTGACGCTCTGCGGATTTAGC 1306
311 ThrLeuCySgInSerGlyIleTyrlLeuValLeuAspIleAlaProAr 327
1307 ACATGTGTCACTCAGTCAGGATTTATATCAATGTCTAGATATGCTCCAG 1356
327 gTySerSerPheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaP 344
1357 GTATTCCAGTTTCTCATGGAGCATCAAGAGATTTTCAGCATAGCAC 1406
344 roValIleValProThrValSerGlyPheLeuLeuSerGlnAspProGlu 360
1407 CTGTCAATGTACCCACGTCTCAGCGATTTCTTCTAGTCAGACCTGAG 1456
361 PheGlyTrpArgAsnValPhePheLeuLeuPheAlaValAsnLeuLeuG 377
1457 TTTGGGTGAGGAATGCTCTCTCTGCTGTTCCGTTAACCTGTTAGG 1506
377 yLeuLeuPheTyrlLeuIlePheGlyGlyAlaAspValGlnGluTrpAla 394
1507 ACTACTCTTCTACCTCATATTGAGAGACAGATGTCAGAAATGGGCTA 1556

```

394 YSGIUAArgLysLeuThrArgLeu 401
|||||
1557 AAGAGAGAAACTACTACTGTTTA 1579

seq_name: gb_pr2:HS283953

seq_documentation_block:

LOCUS HS283953 570 bp mRNA PRI 16-JAN-1997

DEFINITION H.sapiens mRNA; clone CD 274.

ACCESSION Z83953

VERSION Z83953.1 GI:1783342

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..570

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/map="6p22"

/clone="CD 274"

BASE COUNT 127 a 134 c 134 g 171 t 4 others

ORIGIN

alignment_scores:

Quality: 98.00 Length: 98

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x HS283953 ..

Align seg 1/1 to: HS283953 from: 1 to: 570

96 TTPGlyProProGlnGluArgSerArgLeuGlySerLeuAlaLeuSerG 112

|||||

276 TGGGGCCCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 325

|||||

112 YmetLeuGlyCysPheThrAlaLeuLeuGlyGlyPheLeuSerG 129

|||||

326 AATGTTACTGGATGCTTACTGCAATCCATAGAGTGGCTTCAATAGTG 375

|||||

129 LurPheLeuGlyTTPProPheValPheTyrLeuPheGlyGlyValGlyCys 145

|||||

376 AAACCTTGGGGTGGCCCTTGTCTTATATCTTGAAGGTTGGCTGT 425

|||||

146 ValCysCysLeuLeuTTPPheValValIleTyrAspAspProValSerTyr 162

|||||

426 GTCTGCGCTCTCTGCTGTTGTTGATTTATGATGACCCCGTTTCCTA 475

|||||

162 rProTPILeuSerTherSerGluGlyGlyTyrIleIleSerSerLeuGlyG 179

|||||

476 TCCATGCGATAGACACCTCAGAAAAAGAAATACATCATATCCCTTGAAC 525

|||||

179 InGlnValGlySerSerLeuGlnProLeuProIleLeuAlaMet 193

|||||

526 AACAGGTGGGCTCTTCAAGCAGCCTCTCCCATCAAGCTATG 569

seq_name: gb_htg2:AL138726

seq_documentation_block:

LOCUS AL138726 140266 bp DNA HTG 15-FEB-2000

DEFINITION Homo sapiens chromosome 6 clone RP1-139G21, *** SEQUENCING IN

PROGRESS ***, 5 unordered pieces.

ACCESSION AL138726

VERSION AL138726.2 GI:6982716

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (15-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

On Feb 16, 2000 this sequence version replaced gi:691804.

IMPORTANT: This sequence is unfinished and does not necessarily

represent the correct sequence. Work on the sequence is in

progress and the release of this data is based on the understanding

that the sequence may change as work continues. The sequence may

be contaminated with foreign sequence from E.coli, yeast, vector,

phage etc. Order of segments is not known: 800 n's separate

segments. Contig_ID: 00595 length: 20679bp

Contig_ID: 00749 length: 63415bp

Contig_ID: 00902 length: 9637bp

Contig_ID: 01506 length: 27403bp

Contig_ID: 01824 length: 15932bp.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1..20679: contig of 20679 bp in length

* 20680 21479: gap of 800 bp

* 21480 84894: contig of 63415 bp in length

* 84895 85694: gap of 800 bp

* 85695 95331: contig of 9637 bp in length

* 95332 96131: gap of 800 bp

* 96132 123534: contig of 27403 bp in length

* 123535 124334: gap of 800 bp

* 124335 140266: contig of 15932 bp in length.

Location/Qualifiers

source

1..140266

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone="RP1-139G21"

/clone_lib="RPC1-1"

BASE COUNT 41912 a 26597 c 25588 g 42968 t 3201 others

ORIGIN

alignment_scores:

Quality: 71.00 Length: 71

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x AL138726/rev ..

Align seg 1/1 to reverse of: AL138726 from: 1 to: 140266

12 ValProSerLeuGlySerAlaArgTyrGlyIleAlaLeuValLeuHisPhe 28

|||||

65643 GTTCCAGATTATATGTTGCTGCTATGGAATACCCCTGCTTACATTT 65594

|||||

28 eCysAsnPheThrIleAlaGlnAsnValIleMetAsnIleThrMet 45

|||||

```

65593 CTGCAATTCACACAGATAGACAAAATGTCATCATGAAACATCACCATTG 65544
45 AAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
|||||
65543 TAGCCATGTGTACACAGACACAGCCCTCAATCCAGCTCAATGATGATTCCT 65494
62 GluValLeuProValAspSerPheGlyLeuSerLysAlaProLysSe 78
|||||
65493 GAGGTGCTGCTGTGACTCATTTGGTGGCTTAGTAAGCCCAAGAG 65444
78 rleuProAlaLys 82
|||||
65443 TCTTCTCTGCAAG 65431
seq_name: gb_h194:AC012145
seq_documentation_block:
LOCUS AC012145 149155 bp DNA HTG 19-NOV-1999
DEFINITION Homo sapiens clone RP11-11K15, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION AC012145
VERSION AC012145.2 GI:6454004
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 149155)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens, clone RP11-11K15
TITLE Unpublished
COMMENT 2 (bases 1 to 149155)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barina,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deatellano,K., Dewar,K., Domingo,M., Donnellan,L., Doyle,M.,
Gallego,P., Fitzhugh,W., Forrest,C., Funke,K., Gage,D.,
Galligan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGuck,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tittel,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 19, 1999 this sequence version replaced gi:6087932.
All repeats were identified using RepeatMasker:
Smith, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L3182
Center clone name: 11_K_15
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 108943 bases at least Q40
Consensus quality: 127281 bases at least Q30
Consensus quality: 140888 bases at least Q20
Insert size: 149155; sum-of-contrigs
Quality coverage: 4.5 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contrigs. The true order of the pieces

```

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1 1927: contrig of 1927 bp in length
* 1928 gap of unknown length
* 3309 contrig of 1382 bp in length
* 3310 gap of unknown length
* 4771 contrig of 1462 bp in length
* 4772 gap of unknown length
* 5950 contrig of 1179 bp in length
* 5951 gap of unknown length
* 12968 contrig of 7018 bp in length
* 12969 gap of unknown length
* 19842 contrig of 6874 bp in length
* 19843 gap of unknown length
* 29280 contrig of 9438 bp in length
* 45513 gap of unknown length
* 45514 contrig of 16233 bp in length
* 61681 gap of unknown length
* 61680 contrig of 16167 bp in length
* 76539 gap of unknown length
* 76540 contrig of 14859 bp in length
* 96748 gap of unknown length
* 96749 contrig of 20209 bp in length
* 122261 gap of unknown length
* 122262 contrig of 25513 bp in length
* 122262 gap of unknown length
* 122262 149155: contrig of 26894 bp in length.
* Location/Qualifiers
1. 149155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-11K15"
/clone_11b="RPC1-11 Human Male BAC"
BASE COUNT 45294 a 30075 c 29603 g 44177 t 6 others
ORIGIN
1
1 MetGlnValAspGlnThrLeuLeuProAlaLys 11
5269 ATGCAGGTGATGAGACACTGATCCCGAGAAA 5301
-----
alignement_block:
US-09-391-958-1 x AC012145 ..
Align seg 1/1 to: AC012145 from: 1 to: 149155
-----
seq_name: gb_h194:AF182108
seq_documentation_block:
LOCUS AF182108 167065 bp DNA HTG 08-OCT-1999
DEFINITION Homo sapiens chromosome 8 clone BAC R-11m9 map 8p12. 8, ***
SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION AF182108
VERSION AF182108.1 GI:5853325
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 167065)
AUTHORS Schudoy,A., Schlhabel,M., Baumgart,C., Menzel,U., Weber,J.,
Schatteroy,R. and Rosenthal,A.
JOURNAL Submitted (31-AUG-1999) Genome Analysis, Institute of Molecular

```

COMMENT Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
contig 1: pos. 1 - 30930 contig 2: pos. 30931 - 48337 contig 3:
pos. 48338 - 65090 contig 4: pos. 65091 - 92971 contig 5: pos.
92972 - 123996 contig 6: pos. 123997 - 134361 contig 7: pos. 134362
- 134345 contig 8: pos. 134346 - 167062 Indbetween gaps of unknown
size.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

Location/Qualifiers
source 1..167065
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="bp12.8"
/clone="BAC R-11n9"
514..546
/rpt_family="MSR1"
746..770
/note="XPOUND"
/evidence=not_experimental
882..1153
/rpt_family="11m45"
1704..1779
/note="MZEF"
/evidence=not_experimental
1731..1824
/note="SST_REPEAT"
complement(2077..2161)
/rpt_family="11m45"
2079..2156
/rpt_family="11M45"
2080..2210
/rpt_family="11M45A"
complement(2091..2154)
/rpt_family="11m42"
complement(2148..2195)
/rpt_family="11m49"
complement(2148..2195)
/rpt_family="11m45"
2779..2806
/note="SST_REPEAT"
complement(2779..2805)
/note="(CAAAA)n"
/rpt_family="Simple_repeat"
2904..3033
/rpt_family="11m45"
2906..3016
/rpt_family="11"
complement(2916..3181)
/rpt_family="11M4C_Send"
2948..3013
/rpt_family="11m35"
2975..3140
/rpt_type=tandem
3041..3133
/note="SST_REPEAT"
complement(3193..4453)
/rpt_family="11MFC_Send"
3295..3436
/note="SST_REPEAT"
3602..3654
/rpt_family="11m45"
complement(3692..3753)
/note="MZEF"
/evidence=not_experimental
3761..3841
/rpt_family="11m45"
3941..4184
/rpt_family="11m45"
complement(4083..4267)
/note="GENSCAN"

exon complement(4311..4354)
/note="MZEF"
/evidence=not_experimental
4461..4538
/note="XNUN_REPEAT"
4467..4534
/note="SST_REPEAT"
4477..4534
/note="SST_REPEAT"
4478..4533
/rpt_type=tandem
4483..4558
/note="SST_REPEAT"
4485..4534
/note="SST_REPEAT"
4492..4534
/note="SST_REPEAT"
complement(4496..4536)
/note="(TA)n"
/rpt_family="Simple_repeat"
complement(4574..4701)
/note="MZEF"
/evidence=not_experimental
complement(4574..4676)
/note="GENSCAN"
5250..5345
/rpt_family="alu"
5251..5540
/rpt_family="ALU"
5251..5553
/rpt_family="AluSx"
5255..5540
/rpt_family="alu"
5258..5380
/rpt_family="sva"
5259..5319
/rpt_family="sva"
5372..5451
/rpt_family="sva"
5388..5461
/rpt_family="alu"
5405..5464
/rpt_family="sva"
5458..5512
/rpt_family="sva"
5529..5562
/note="XNUN_REPEAT"
5533..5562
/note="SST_REPEAT"
5533..5645
/note="SST_REPEAT"
5533..5553
/note="SST_REPEAT"
complement(6176..6264)
/note="MZEF"
/evidence=not_experimental
complement(6176..6336)
/note="GRAIL"
/evidence=not_experimental
complement(6425..6488)
/note="MZEF"
/evidence=not_experimental
6482..6548
/note="SST_REPEAT"
6624..6686
/note="SST_REPEAT"
6624..6658
/note="SST_REPEAT"
6624..6698
/note="XNUN_REPEAT"
6630..6698
/note="SST_REPEAT"
6639..6675

```

exon                                     /note="SST_REPEAT"
6727..6814
/note="MZEF"
/evidence=not_experimental
exon                                     /note="GRAIL"
6755..6814
/evidence=not_experimental
repeat_region complement(7342..7506)
repeat_region /rpt_family="MIR"
8446..8609
/note="SST_REPEAT"
complement(8489..8518)
exon                                     /note="MZEF"
/evidence=not_experimental
repeat_region complement(8615..8769)
repeat_region /rpt_family="11pa1"
8683..8767
/rpt_family="11pa2"
complement(8683..8767)
repeat_region /rpt_family="11pa2"
8684..8765
/rpt_family="11pa5"
complement(8687..8767)
repeat_region /rpt_family="11pa7"
complement(8787..9554)
repeat_region /rpt_family="LIM4orf2"
9077..9126
/note="XMUN_REPEAT"
complement(9080..9400)
repeat_region /rpt_family="Aluub"
9080..9126
/note="SST_REPEAT"
complement(9122..9281)
repeat_region /rpt_family="Alu"
complement(9126..9287)
repeat_region /rpt_family="Alu"
complement(9136..9195)
repeat_region /rpt_family="Sva"
complement(9178..9260)

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AF182108/rev ..
Align seg 1/1 to reverse of: AF182108 from: 1 to: 167065

276 ILEuGlySerLeuProSerSerAlaLeu 285
|||||
160806 ATTTGGGTCGTACTCTTCACGCTCA 160777

seq_name: gb_p12:CNS019DX

seq_documentation_block:
LOCUS CNS019DX 660 bp mRNA PIN 02-SEP-1999
DEFINITION Botrytis cinerea strain T4 cDNA library under conditions of
nitrogen deprivation.
ACCESSION AF111517.1 GI:5826136
VERSION cDNA library; nitrogen deprivation.
KEYWORDS Botryotinia fuckeliana.
SOURCE Botryotinia fuckeliana
ORGANISM Eukaryota; Fungi; Ascomycota; Euascomycetes; Discomycetes;
Leotiales; Sclerotiniaceae; Botryotinia.
REFERENCE 1 (bases 1 to 660)
AUTHORS Bitton,F., Lewis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
78026 Versailles, France
REFERENCE 2 (bases 1 to 660)

```

```

AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage ;
CP 5706 91057 Evry cedex FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT The cDNA library to be analyzed within the framework of this
project was created using a Botrytis cinerea strain which was grown
under conditions of nitrogen deprivation, which is the normal
situation for B. cinerea during its development on its host plant.
The library was produced in an oriented direction, in the pBSII
vector.

FEATURES
source Location/Qualifiers
1..660
/organism="Botryotinia fuckeliana"
/strain="T4"
/db_xref="taxon:40559"
/note="Genoscope sequence ID : W37C021"
BASE COUNT 157 a 152 c 148 g 203 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x CNS019DX ..
Align seg 1/1 to: CNS019DX from: 1 to: 660

348 ProThrvAlserGlyPheLeuLeuSer 356
|||||
442 CCCACAGCTCAGGCTCTCTTTCACA 468

seq_name: gb_p12:DPCPNAM

seq_documentation_block:
LOCUS DPCPNAM 865 bp mRNA PIN 30-SEP-1999
DEFINITION D. carota mRNA for proliferating cell nuclear antigen (PCNA).
ACCESSION X62977.1 GI:18361
VERSION X62977.1 GI:18361
KEYWORDS Proliferating cell nuclear antigen.
SOURCE carrot.
ORGANISM Daucus carota
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
REFERENCE 1 (bases 1 to 865)
AUTHORS Hata,S., Kouchi,H., Tanaka,Y., Minami,E., Matsumoto,T., Suzuki,I.
and Hashimoto,J.
TITLE Identification of carrot cDNA clones encoding a second putative
proliferating cell-nuclear antigen, DNA polymerase delta auxiliary
protein
JOURNAL Eur. J. Biochem. 203 (3), 367-371 (1992)
MEDLINE 92137220
REFERENCE 2 (bases 1 to 865)
AUTHORS Hata,S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1991) S.Hata, Faculty of Science, Himeji
Institute of Technology, Kamigori, Ako-gun, Hyogo 678-12, Japan
FEATURES
source Location/Qualifiers
1..865
/organism="Daucus carota"
/cultivar="Kurodagosun"
/db_xref="taxon:4039"
BASE COUNT 245 a 168 c 222 g 230 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0

```

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x DCPNAM/rev ..

Align seg 1/1 to reverse of: DCPNAM from: 1 to: 865

279 SerleupProserSerAlaLeuIleVal 287
|||||
295 TCACCTGCATCATCAGCCTTAATAGTG 269

seq_name: gb_p12:AF202292

seq_documentation_block:

LOCUS AF202292 870 bp DNA PLN 27-DEC-1999
DEFINITION Euscomycete sp. strain CI 3-10 18S ribosomal RNA gene, partial
sequence.

ACCESSION AF202292
VERSION AF202292.1 GI:6636194

KEYWORDS
SOURCE Euscomycete sp.

ORGANISM Euscomycete sp.
Eukaryota; Fungi; Ascomycota.

REFERENCE 1 (bases 1 to 870)
Vandenkoornhuyse, P., Martel, M. H. and Leyval, C.

AUTHORS
TITLE Global fungal diversity associated to the roots of Arrhenatherum
elatus

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 870)
Vandenkoornhuyse, P., Martel, M. H. and Leyval, C.

AUTHORS
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) CPB, CNRS UPR6831, 17, rue Notre-Dame des
Pauvres, Vandœuvre-lès-Nancy F-54500, France

FEATURES
source
1..870
/organism="Euscomycete sp."
/strain="CI 3-10"
/db_xref="taxon:110117"
<1..>870
/product="18S ribosomal RNA"

RNA
/product="18S ribosomal RNA"

BASE COUNT 236 a 175 c 241 g 218 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x AF202292/rev ..

Align seg 1/1 to reverse of: AF202292 from: 1 to: 870

6 ThirleupProArqlyValProser 14
|||||
743 ACTTGATTCCTCGTAGGTCGCGAGC 717

seq_name: gb_p11:NTA18135

seq_documentation_block:

LOCUS NTA18135 922 bp mRNA PLN 02-NOV-1998
DEFINITION Nicotiana tabacum mRNA for PCNA1 protein.
ACCESSION Y18135
VERSION Y18135.1 GI:3821258
KEYWORDS PCNA1 gene.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asterales; Solanales; Solanaceae; Nicotiana.

REFERENCE 1 (bases 1 to 922)
Duong van, H., Gaikwad, A.S. and Mukherjee, S.

AUTHORS

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 922)
AUTHORS Mukherjee, S.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1998) S. Mukherjee, International Center for
Genetic, Engineering And Biotechnology, Plant Molecular Biology
Lab, Po Box: 10504, Aruna Asaf Ali Marg, New Delhi 110067, INDIA

FEATURES
source
1..922
/organism="Nicotiana tabacum"
/db_xref="taxon:4097"
/tissue_type="leaf"
1..39
/gene="PCNA1"
1..922
/gene="PCNA1"
40..840
/gene="PCNA1"
/product="PCNA protein"
/protein_id="CAA77062.1"
/db_xref="GI:3821259"
/translation="MDELRLVGSLLKRVLESIKELVNDANFDCSSGTFSLQAMDSH
VALVALLRSEGEHYRCDBRNLSMGMLNNMAKMKLCAANDIITIKADGSDVTFM
FSPPODKSIDSEFKIMIDSEHLGIPEDYKAIYIMPAEFARICKDLSISGDTYVI
AVSKGVKSTSGDIDGSANIVROMTVDKPEATYIENNEPVALQFALRYNNSFKA
TPLSSTVITSLSNELPVYEVKIAEMGVRYFLAKRIEDEETKPA"
841..922
/gene="PCNA1"

5'UTR
gene
CDS

3'UTR

BASE COUNT 246 a 177 c 227 g 272 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x NTA18135/rev ..

Align seg 1/1 to reverse of: NTA18135 from: 1 to: 922

279 SerleupProserSerAlaLeuIleVal 287
|||||
329 TCACCTGCATCATCAGCCTTAATAGTA 303

seq_name: gb_p11:AB008186

seq_documentation_block:

LOCUS AB008186 948 bp mRNA PLN 17-SEP-1998
DEFINITION Pisum sativum mRNA for proliferating cell nuclear antigen, complete
cds.
ACCESSION AB008186
VERSION AB008186.1 GI:3608174
KEYWORDS proliferating cell nuclear antigen.
SOURCE Pisum sativum cDNA to mRNA.
ORGANISM Pisum sativum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Fabales; Fabaceae; Papilionoideae; Pisum.

REFERENCE 1 (bases 1 to 948)
Shimizu, S. and Mori, H.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1997) to the DDBJ/EMBL/GenBank databases. Hitoshi
Mori, Nagoya University, School of Agricultural Sciences, Furo-cho
Chikusa-ku, Nagoya, Aichi 464-01, Japan
(E-mail: mori@tohoku.ac.jp, nagoya-u.ac.jp, Tel: 052-789-4167,
Fax: 052-789-4296)
2 (sites)
Analysis of cycles of dormancy and growth in pea axillary buds
based on mRNA accumulation patterns of cell cycle-related genes

AUTHORS
TITLE

JOURNAL Plant Cell Physiol. 39 (3), 255-262 (1998)
MEDLINE 98249622
FEATURES
Source Location/Qualifiers
1..948
/organism="Pisum sativum"
/db_xref="taxon:3888"
CDS
2..802
/function="auxiliary protein of DNA polymerase-delta"
/codon_start=1
/product="proliferating cell nuclear antigen"
/protein_id="BAA3351.1"
/db_xref="GI:3608175"
/translation="MLELRVQSLKLVLESIKELVNDANFDCSSTGFSIQAMDSH
VALVALLRSEGFHEHYRCDRLNSGMNINNAKMLKACAGNDITIKADGSDVYTM
FESPTQDKISDFEMKIMIDISEHLGIPAEYHAIVRMPSEAFRICDLSIGDTVVI
AVSKGVKSTGKIDIGSANIVYCRONTYDCKPEATVEMNEPVLAQLRMRNSFTKA
TPLSSVTSISLSNELPYVVEYKIAEMGVRYRLAPKIEDEETKPOA"
BASE COUNT 251 a 179 c 225 g 293 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x AB008186/rev ..

Align seg 1/1 to reverse of: AB008186 from: 1 to: 948

279 SerleupProSerSerAlaLeuIleVal 287
|||||
291 TCACCTCCATCATCAGCCTTAATAGTA 265

seq_name: gb_p11:PSPCNA

seq_documentation_block:

LOCUS PSPCNA 990 bp mRNA PLN 11-FEB-1999
DEFINITION Pisum sativum mRNA for proliferating cell nuclear antigen, complete
CDS.
Y16796
Y16796.1 GI:3392983
KEYWORDS
PCNA gene; proliferating cell nuclear antigen.
SOURCE
pea.
ORGANISM
Pisum sativum

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Fabales; Fabaceae; Papilionoideae; Pisum.
1 (bases 1 to 990)
Duong van H., Gaikwad, A.S. and Mukherjee, S.K.
JOURNAL
Unpublished
2 (bases 1 to 990)
Mukherjee, S.
Direct Submission
Submitted (04-MAR-1998) S. Mukherjee, International Center for
Genetic Engineering and Biotechnology, Plant Molecular Biology
Lab, PO Box 10504, Aruna Asaf Ali Marg New Delhi 110067, INDIA
Location/Qualifiers
1..990
/organism="Pisum sativum"
/db_xref="taxon:3888"
/clone_lib="lambda Uni-Zap"
1..11
/gene="PCNA"
1..990
/gene="PCNA"
12..812
/gene="PCNA"
/codon_start=1
/product="proliferating cell nuclear antigen"
/protein_id="CAA76392.1"
/db_xref="GI:4379382"

REFERENCE
AUTHORS
JOURNAL
TITLE
MEDLINE
COMMENT
On Mar 21, 1997 this sequence version replaced gi:1877023.
Sequence updated (18-Mar-1997) by: Kunio Ihara.
Location/Qualifiers
1..1467
/organism="Haloarcula vallismortis"
/db_xref="taxon:28442"
1..1467
/product="16S rRNA"
BASE COUNT 351 a 376 c 472 g 268 t
ORIGIN

FEATURES
Source Location/Qualifiers
1..1467
/organism="Haloarcula vallismortis"
/db_xref="taxon:28442"
1..1467
/product="16S rRNA"
BASE COUNT 351 a 376 c 472 g 268 t
ORIGIN

5'UTR
gene
CDS

/translation="MLELRVQSLKLVLESIKELVNDANFDCSSTGFSIQAMDSH
VALVALLRSEGFHEHYRCDRLNSGMNINNAKMLKACAGNDITIKADGSDVYTM
FESPTQDKISDFEMKIMIDISEHLGIPAEYHAIVRMPSEAFRICDLSIGDTVVI
AVSKGVKSTGKIDIGSANIVYCRONTYDCKPEATVEMNEPVLAQLRMRNSFTKA
TPLSSVTSISLSNELPYVVEYKIAEMGVRYRLAPKIEDEETKPOA"
3'UTR
BASE COUNT 283 a 181 c 229 g 297 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x PSPCNA/rev ..

Align seg 1/1 to reverse of: PSPCNA from: 1 to: 990

279 SerleupProSerSerAlaLeuIleVal 287
|||||
301 TCACCTCCATCATCAGCCTTAATAGTA 275

seq_name: gb_bal:D50851

seq_documentation_block:

LOCUS D50851 1467 bp DNA BCT 10-FEB-1999
DEFINITION Haloarcula vallismortis DNA for 16S rRNA.
ACCESSION D50851
Y16796
Y16796.1 GI:1902934
KEYWORDS
16S ribosomal RNA.
SOURCE
Haloarcula vallismortis
Haloarcula vallismortis
Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
Haloarcula.
1 (bases 1 to 1467)
Ihara, K.
Direct Submission
Submitted (02-JUN-1995) to the DDBJ/EMBL/GenBank databases. Kunio
Ihara, Nagoya University, School of Science, Department of Biology;
Furo-cho, Chikusa-ku, Nagoya, Aichi 464-01, Japan
(E-mail: i45031a@nuc.cc.nagoya-u.ac.jp, Tel: 052-789-2970,
Fax: 052-789-2968)
2 (bases 1 to 1467)
Ihara, K., Watanabe, S. and Mukohata, Y.
Haloarcula argentinensis spec. nov. and Haloarcula mukohatae sp.
nov., two new species of extreme halophilic archaea from argentina
3 (sites)
Unpublished (1995)
JOURNAL
TITLE
AUTHORS
Ihara, K., Watanabe, S. and Tamura, T.
Haloarcula argentinensis sp. nov. and Haloarcula mukohatae sp.
nov., two new extremely halophilic archaea collected in Argentina
Int. J. Syst. Bacteriol. 47 (1), 73-77 (1997)
J97148980
COMMENT
On Mar 21, 1997 this sequence version replaced gi:1877023.
Sequence updated (18-Mar-1997) by: Kunio Ihara.
Location/Qualifiers
1..1467
/organism="Haloarcula vallismortis"
/db_xref="taxon:28442"
1..1467
/product="16S rRNA"
BASE COUNT 351 a 376 c 472 g 268 t
ORIGIN

REFERENCE
AUTHORS
JOURNAL
TITLE
MEDLINE
COMMENT
On Mar 21, 1997 this sequence version replaced gi:1877023.
Sequence updated (18-Mar-1997) by: Kunio Ihara.
Location/Qualifiers
1..1467
/organism="Haloarcula vallismortis"
/db_xref="taxon:28442"
1..1467
/product="16S rRNA"
BASE COUNT 351 a 376 c 472 g 268 t
ORIGIN

FEATURES
Source Location/Qualifiers
1..1467
/organism="Haloarcula vallismortis"
/db_xref="taxon:28442"
1..1467
/product="16S rRNA"
BASE COUNT 351 a 376 c 472 g 268 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x D50851/rev ..

Align seg 1/1 to reverse of: D50851 from: 1 to: 1467

349 ThrValSerGlyPheLeuLeuSerGln 357
|||||
298 ACCGTCGCCGGGTTCTGCTCTCACAA 272

seq_name: gb_ba1:AB010965

seq_documentation_block:

LOCUS AB010965 1470 bp DNA BCT 03-AUG-1999
DEFINITION Halocaula sp. gene for 16S rRNA.
ACCESSION AB010965
VERSION AB010965.1 GI:4115513
KEYWORDS 16S rRNA; 16S ribosomal RNA.
SOURCE Halocaula sp. (strain:801030/1) DNA.
ORGANISM Halocaula sp.
Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;

REFERENCE
AUTHORS 1 (sites)
TITLE Oren, A., Ventosa, A., Gutierrez, M.C. and Kamekura, M.
Halocaula quadrata sp. nov., a square, motile archaean isolated from a brine pool in Sinai
Int. J. Syst. Bacteriol. 49 Pt 3, 1149-1155 (1999)

JOURNAL 99354505
REFERENCE 2 (bases 1 to 1470)
AUTHORS Kamekura, M.
TITLE Direct Submission
SUBMITTED (09-FEB-1998) to the DDBJ/EMBL/GenBank databases.
Masahito Kamekura, Noda Institute for Scientific Research, 399
Noda, Chiba 278, Japan (E-mail: kamekura@super.nig.ac.jp,
Tel: 0471-23-5573, Fax: 0471-23-5959)

FEATURES
source
1..1470
/organism="Halocaula sp."
/strain="801030/1"
/db_xref="taxon:44098"
1..1470
/product="16S rRNA"

rRNA

BASE COUNT 350 a 381 c 476 g 263 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AB010965/rev ..

Align seg 1/1 to reverse of: AB010965 from: 1 to: 1470

349 ThrValSerGlyPheLeuLeuSerGln 357
|||||
298 ACCGTCGCCGGGTTCTGCTCTCACAA 272

seq_name: gb_ba1:D63786

seq_documentation_block:

LOCUS D63786 1472 bp DNA BCT 07-AUG-1997
DEFINITION Halorubrum trapanicum DNA for 16S rRNA.
ACCESSION D63786
VERSION D63786.1 GI:2315141
KEYWORDS 16S ribosomal RNA.
SOURCE Halorubrum trapanicum (strain:JCM8979) DNA.
ORGANISM Halorubrum trapanicum
Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;

REFERENCE
AUTHORS 1 (bases 1 to 1472)
Kamekura, M.

TITLE Direct Submission
JOURNAL Submitted (06-AUG-1995) to the DDBJ/EMBL/GenBank databases.
Masahito Kamekura, Noda Institute for Scientific Research, 399
Noda, Chiba 278, Japan (E-mail: kamekura@super.nig.ac.jp,
Tel: 0471-23-5573, Fax: 0471-23-5959)

REFERENCE
AUTHORS 2 (sites)
TITLE Kamekura, M., Dyall-Smith, M.L., Upasani, V., Ventosa, A. and Kates, M.
Diversity of alkaliphilic halobacteria: proposals for transfer of
Natronobacterium vacuolatum, Natronobacterium magadii, and
Natronomonas gen. nov., respectively, as Halorubrum vacuolatum
comb. nov., Natrhalba magadii comb. nov., and Natronomonas
pharaonis comb. nov., respectively
Int. J. Syst. Bacteriol. 47 (3), 853-857 (1997)

JOURNAL 97370612
MEDLINE Location/Qualifiers
FEATURES
1..1472
/organism="Halorubrum trapanicum"
/strain="JCM8979"
/db_xref="taxon:29284"
1..1472

rRNA

BASE COUNT 340 a 389 c 479 g 264 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x D63786/rev ..

Align seg 1/1 to reverse of: D63786 from: 1 to: 1472

349 ThrValSerGlyPheLeuLeuSerGln 357
|||||
298 ACCGTCGCCGGGTTCTGCTCTCACAA 272

OM of: US-09-391-958-1 to: N_Geneseq_36:* out_format: pfs
Date: May 25, 2000 3:28 PM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framet-p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09391958/runtat.24052000.213606.8888/app-query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=oligo.rng -GAP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAP=4.500 -OGAPEXT=0.050 -XGAP=60.000 -XGAPEXT=60.000
-FGAP=6.000 -FGAPEXT=7.000 -YGAP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human0.ccl -LIST=45 -DOCLIST=200 -THR.SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=100000 -USER=US09391958 -NCPU=6 -ICPU=3 -NO_XLUPY -WAIT
-THREADS=1

Search information block:
Query: US-09-391-958-1
Query length: 401
Database: N_Geneseq_36:*
Database sequences: 31585
Database length: 125096042
Search time (sec): 62.170000

WARN: XGAP and YGAP must be equal. Assuming YGAP=XGAP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:V43711	+	401.00	7157.01	0.0	1643	Human sodium-dependent phosphatase
N_Geneseq_36:V57910	+	241.00	4289.10	2.6e-231	1795	Human haemochromatosis gene NT
N_Geneseq_36:V43712	+	85.00	1507.49	2.3e-76	272	Sodium-dependent phosphate cotr
N_Geneseq_36:V57909	+	9.00	129.81	12.4e	2281	Human haemochromatosis gene NT
N_Geneseq_36:V57926	+	9.00	95.51	1.0e+03	235033	Hereditary haemochromatosis
N_Geneseq_36:V57903	+	9.00	95.43	1.0e+03	237326	Hereditary haemochromatosis
N_Geneseq_36:V43714	+	8.00	145.29	1.71	25	Primer for Sodium-dependent phos
N_Geneseq_36:Q14339	+	8.00	113.73	98.23	1779	Human epithelial precursor. New
N_Geneseq_36:V82825	+	8.00	112.52	114.72	2095	Human GP88 autoocrine growth fa
N_Geneseq_36:V82825	+	8.00	112.42	116.22	2124	Granulin coding sequence. New
N_Geneseq_36:Q49052	+	8.00	112.16	120.06	2198	Granulin coding sequence. New
N_Geneseq_36:Q70754	+	8.00	112.14	120.48	2206	Beta tubulin gene of Acremoniu
N_Geneseq_36:Q48230	+	8.00	108.84	183.93	3445	Acremonium chrysogenum beta-tu
N_Geneseq_36:Q54055	+	8.00	108.84	183.93	3445	A. chrysogenum beta-tubulin 116
N_Geneseq_36:Q5406	+	8.00	108.84	183.93	3445	A. chrysogenum beta-tubulin 116
N_Geneseq_36:V43713	+	7.00	127.37	17.07	25	Primer for Sodium-dependent phos
N_Geneseq_36:V75542	+	7.00	123.89	26.67	40	Streptococcus pneumoniae ORF clc
N_Geneseq_36:T19786	+	7.00	117.57	60.01	94	Human gene signature HUMG500862.
N_Geneseq_36:X12674	+	7.00	111.59	129.29	211	Human biallelic polymorphic DNA
N_Geneseq_36:X12675	+	7.00	111.59	129.29	211	Human biallelic polymorphic DNA
N_Geneseq_36:Q59355	+	7.00	111.42	137.42	225	Human brain Expressed Sequence
N_Geneseq_36:T75168	+	7.00	110.24	153.60	253	Sequence used in development of
N_Geneseq_36:Q68845	+	7.00	110.07	157.05	259	Elastase inhibiting peptide DNA
N_Geneseq_36:Q26065	+	7.00	109.47	170.26	282	21B4 gene clone pT13. EcoRI in
N_Geneseq_36:V16717	+	7.00	109.06	178.85	297	Nucleotide sequence encoding th
N_Geneseq_36:V89417	+	7.00	108.91	182.28	303	EST clone CM532. New polynucle
N_Geneseq_36:V20230	+	7.00	108.79	185.13	308	EST clone CM532. New polynucle
N_Geneseq_36:V87628	+	7.00	108.08	202.77	339	EST clone BV167. New polynucle
N_Geneseq_36:Q79466	+	7.00	107.97	205.61	344	Mouse T118 Flt3 ligand peptide
N_Geneseq_36:Q78504	+	7.00	107.80	210.15	352	Phosphate starvation-induced pr
N_Geneseq_36:T76576	+	7.00	107.59	215.61	362	Human gene signature HUMS08822
N_Geneseq_36:V35856	+	7.00	107.39	221.46	372	Human encoding protein homologu
N_Geneseq_36:V9553	+	7.00	107.33	221.06	372	Arabidopsis lysine ketoglutarat
N_Geneseq_36:V87095	+	7.00	106.93	235.46	396	EST clone BB222. New polynucle
N_Geneseq_36:X18934	+	7.00	106.87	236.69	399	EED2 DNA sequence. Identifying
N_Geneseq_36:N81136	+	7.00	106.82	238.38	402	Heaetoxycarphaeospoirtin-C-synt
N_Geneseq_36:Q59973	+	7.00	106.74	240.63	406	Human brain Expressed Sequence
N_Geneseq_36:T74719	+	7.00	105.80	271.47	461	Human papillomavirus type 53 L1
N_Geneseq_36:T78155	+	7.00	105.80	271.47	461	DNA sequence of L1 amplified rd

N_Geneseq_36:X13446	+	7.00	105.23	292.11	498	Enterococcus faecalis genome
N_Geneseq_36:T31834	+	7.00	104.93	303.79	519	Human papilloma virus E6/E7
N_Geneseq_36:T87771	+	7.00	104.34	327.63	562	EST clone BQ261. New polynuc
N_Geneseq_36:V88831	+	7.00	104.30	329.29	565	EST clone HR560. New polynuc
N_Geneseq_36:V53093	+	7.00	104.22	332.61	571	DNA encoding a S. pneumoniae
N_Geneseq_36:T75153	+	7.00	103.56	361.85	624	Metastasis inhibitor cDNA. C

seq_name: N_Geneseq_36:V43711
seq_documentation_block:
ID V43711 standard; DNA. 1643 BP.
AC V43711.
DT 30-NOV-1998 (first entry)
DE Human sodium-dependent phosphate cotransporter coding sequence.
KW Sodium-dependent phosphate cotransporter; human; NAFTR; cancer; myopathy;
KW cell signalling disorder; phosphate regulation disorder; therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 235..1440
FT /tag= a

MO9837198-A1.
PD 27-AUG-1998.
PF 24-FEB-1998: U03745.
PR 24-FEB-1997: US-805118.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Lai P.
DR WPI: 98-467561/40.
DR P-PSDB: W69971.
PT New isolated human sodium-dependent phosphate co:transporter - used
PT to develop products for treating e.g. cancers, osteoporosis,
PT Alzheimer's disease, diabetes, encephalopathy, myopathy,
PT hypocalcaemia or hypoglycaemia.
PS Claim 5: Fig 1: 66pp: English.
CC This sequence encodes the human sodium-dependent phosphate cotransporter
(NAFTR) of the invention. NAFTR and agonists of it can be used to treat
CC or prevent disorders associated with decreased phosphate levels,
CC e.g. cancers of the kidney, disorders of decreased phosphate levels
CC including tumoral calcinosis, osteomalacia, osteoporosis, familial
CC hypophosphataemia, rickets, cystineuria, nephrocalcinosis,
CC glomerulonephritis, renal calculus, Alzheimer's disease, diabetes
CC mellitus, hereditary amyloidosis, myopathies including progressive
CC external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic epilepsy,
CC encephalopathy, cardiomyopathy, hypokalaemia, Goodpasture's syndrome, and
CC disorders of cell signalling through cAMP, ATP, NADPH and
CC glucose-6-phosphate. Antagonists or inhibitors of NAFTR may be
CC administered to a subject to treat or prevent disorders associated with
CC increased phosphate levels, e.g. hypocalcaemia, hypocalcaemia, and
CC abnormal phosphate regulation in neurons, gastrointestinal tract and
CC liver. The products can also be used for detection, diagnosis and drug
CC screening.

Sequence	1643 BP;	435 A;	376 C;	359 G;	473 T;
----------	----------	--------	--------	--------	--------

alignment_scores:
Quality: 401.00 Length: 401
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x V43711 ..
Align seg 1/1 to: V43711 from: 1 to: 1643

1	MeGlnValAspGluThrIleProArgLysValProSerLeuGlyCysE	17
235	ATCCAAATGATGATGACACGATCCACGAAAGTTCCAAATTATCTTC	284
17	raLaarGTYrGlyIleAlaLeuValLeuHisPheCysAspPheThrI	34
285	TGCTTCGTAATGAAATACCTCGTCTTACATTTCTGCAATTTCAACA	334
34	leAlaGlnAsValIleMetAsnIleThrMetValAlaMetValAsnSer	50

```

335 TAGCAAAAATGTCATCATGAACATCACCATGATGACCATGGTCAACAGC 384
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGlnValLeuProValAs 67
385 ACAAGCCCTCAATCCACCTCATATATCTCTGAGGGTGGCTGTTGA 434
67 pSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSers 84
435 CTCAATTGGTGGCCCTAAAGTAAAGCCCAAGAGATCTCTCCCAAGTCTC 484
84 erilleLeuGlyGlyGlnPheAlaIleIleProGlnValGlyProGln 100
485 CAATACCTGGGGGTGAGTTTGCATTTGGGAAAGTGGGGCCCTCCCA 534
101 GluArgSerArgLeuGlySerLysAlaLeuSerGlyMetLeuLeuGly 117
535 GAACGAAGACAGACTGAGCATGGCTTATACAGAAATGTTACTGGGATG 584
117 sPheThrAlaIleLeuIleGlyGlyPheIleSerGlyThrLeuGlyTrp 134
585 CTTTACTGGCCATCCCTCATAGTGGCTTCATTTAGAAACCTGGGTGGC 634
134 ropPheValPheTyrIlePheGlyGlyValGlyCysValGlyCysLeu 150
635 CCTTGTCTCTATATCTTTGGAGGTGTGGCTGTCTCTGCTCTCTCTC 684
151 TrpPheValIleTyrAspAspProValSerTyrProTrpIleSerTh 167
685 TGGTTGTGTGATTTATGATGACCCCGTTCTCATCCAGATGATTAAGC 734
167 rSerGlyLysGlyThrIleIleSerSerLeuLysGlnGlnValGlySers 184
735 CTGAGAAAAAGAAATACATCATCTCTGTTGAACAAACAGTGGGTCTT 784
184 erLysGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTrp 200
785 CTAAAGAGCCTTCCCATCAAAAGCATATGCTCATCCATTTGG 834
201 SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThMet 217
835 TCCATATGTTTGGCTGTGACCCCTTCATATGATGATGACCAATGTG 884
217 ValTyrIleProThrTyrIleSerSerValIleHisValAsnIleArg 234
885 TGTATACATACCACTTACATCACTGTGTGACCAATGTTAACACAGAG 934
234 sPAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
935 ACAATGAGACTTCTATCTGCCCTTCTTATTTGTTGCCGTGATAGGC 984
251 MetValGlyGlyTyrLeuAlaAspPheLeuThrLysLysPheArgLe 267
985 ATGGTGGAGGCTATCTGGCAGATTTCTTAAACCAAAAGTTTAACT 1034
267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSera 284
1035 CATCACTGAGGAGAAATTTGCCACAATTTAGAAAGTCTCCCTTTAG 1084
284 lalLeuIleValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThr 300
1085 CACTATTGTGTCTCTGCCCTTACCTCAATTCGGGTATATACAGCAACT 1134
301 AlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuGlySerglyTr 317
1135 GCCTTGCTGAGCGCTCTCTGGGAGATTTAGCAGATTTGTCTAGGAT 1184
317 eTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMet 334
1185 TTATATCAATGCTTATGATATTTGCCCAAGATATTCAGTTTCTCATGG 1234
334 lYAlaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
1235 GAGCATCAAGAGGATTTTTCAGCATACACCTGTCATTTGACCACTGTC 1284

```

```

351 SerGlyPheLeuLeuSerGlnAspProGlnPheGlyTrpArgAsnValP 367
1285 AGCGGATTTCTTCTTACTACAGACCTGATTTGGGTGGAGAAATGCTT 1334
367 ePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuIle 384
1335 CTCTTCTCTTTTGGCCCTTAACCTGTAGGACTACTCTTCTACTCATAT 1384
384 heGlyGlnAlaAspValGlnGlnTrpAlaLysGlnArgLysLeuThrArg 400
1385 TTGGAGAGACAGATGTCACAAATGGCTTAAAGAGAGAAAACTACTCGT 1434
401 Leu 401
1435 TTTA 1437

seq_name: N_Geneseq_36:V57910
seq_documentation_block:
ID V57910 standard; cDNA; 1795 BP.
AC V57910;
DT 21-DEC-1998 (first entry)
DE Human haemochromatosis gene NTP4.
KW Bovine butyrophilin; Bf; human hereditary haemochromatosis; HFE;
diagnosis; Iron metabolism; NPT3; NPT4; Rofet; BFP1; BFP2; BFP3;
KW BFP4; BFP5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 377..1582
FT FT /*tag= a
FT ET /product= "NTP4"
FN WO9814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Krommal GS, lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
DR WPI: 98-240014/21.
DR P-PSDB: W78920.
PT Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 58; Fig 7; 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC HFE gene sequences from the present invention in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BFP genes, which are homologues of the milk
CC protein butyrophilin (Bf), and can be used in the production of agonists
CC and antagonists of Bf function. Also described are: (1) a Rofet gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia. The present sequence encodes
SQ Sequence 1795 BP; 473 A; 436 C; 402 G; 484 T;

alignment_scores:
Quality: 241.00 Length: 241
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:
US-09-391-958-1 x V57910 ..

Align seg 1/1 to: V57910 from: 1 to: 1795

```
161 SerTyrProTrpIleSerThrSerGluIysGluTyrIleIleSerSerle 177
|||||
857 TCCTATCCATGATTAAGACCTCAGAAAGAAATACATCATATCCTCTT 906
|||||
177 ULVSGINGINValGlySerSerIysGlnProLeuProIleLysAlaMetL 194
|||||
907 GAAACAAACAGCTGGGCTCTTAAAGCAGCCTTCCATCAAGCTATGTC 956
|||||
194 euTrpSerLeuProIleTrpSerIleCysLeuGlyCysPheSerHisGln 210
|||||
957 TCGATCTCTACCATTTGGTCCATATGTTAGGCTGTTTCAGCATCAAA 1006
|||||
211 TrpLeuValSerThrMetValValTyrIleProThrTyrIleSerSerVa 227
|||||
1007 TGGTTAGTTAGCAACAATGGTTGATATACATCAACTTACATGCTCTGT 1056
|||||
227 LTYRHSValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuProPheI 244
|||||
1057 GTACCATGTTACATCAGACAGACAATGCACTTATCTGCCCTTCTTTTA 1106
|||||
244 LeValAlaTrpValIleGlyMetValGlyGlyTyrLeuAlaAspPheLeu 260
|||||
1107 TTGTGGCTGGGTATAGCATGGTGGAGGCTATCTGSCAGATTTCTT 1156
|||||
261 LeuThrLysLysPheArgLeuIleThrValArgLysIleAlaThrIleLe 277
|||||
1157 CTAAACCAAAAGTTTAGACTCATCAGCTGTGAGAAATTTGCCACAATTTT 1206
|||||
277 uGlySerLeuProSerSerAlaLeuIleValSerLeuProTyrLeuAsn 294
|||||
1207 AGGAAGTCTCCCTCTCAGACACATGTTGTCTCTGCTTACCTCAAT 1256
|||||
294 erGlyTyrIleThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSer 310
|||||
1257 CCGGCTATATCAGCAAGTCCCTGCTGACGCTCTTCCGGATTATAGC 1306
|||||
311 ThrLeuCysGlnSerGlyIleTyrIleAsnValLeuAspIleAlaProAr 327
|||||
1307 ACATTTGCTCAGTCAGGAGATTATATCAATGTCTTATATATTTGCTCAAG 1356
|||||
327 gTyrSerSerPheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaP 344
|||||
1357 GFAATCCAGTTTCTCTCATGGAGCATCAAGAGATTTTCGAGCATAGCAG 1406
|||||
344 roValIleValProThrValSerGlyPheLeuLeuSerGlnAspProGlu 360
|||||
1407 CTGTCAATGTAGCCACATGTCAGCGGATTTCTTTAGTCAGAGCCCTGAG 1456
|||||
361 PheGlyTrpArgAsnValPhePheLeuLeuPheAlaValAsnLeuLeuG 377
|||||
1457 TTTGGGTGGAGGAATGCTCTTCTCTGTTTCCCTTAACCTGTTAGG 1506
|||||
377 yLeuLeuPheTyrLeuIlePheGlyGlyAlaAspValGlnGluTrpAlaL 394
|||||
1507 ACTACTCTTCTACCTCATATTTGGAGAAAGCAGATGTCCAAGAAATGGGCTA 1556
|||||
394 ySGluArgGlyLeuThrArgLeu 401
|||||
1557 AAGAGAGAAAACCTCATCTCGTTTA 1579
|||||
```

seq_name: N_Geneseq_36:V43712

seq_documentation_block:

ID V43712 standard; DNA: 272 BP.
AC V43712;
DT 30-NOV-1998 (first entry)
DE Sodium-dependent phosphate cotransporter coding sequence fragment.
KW Sodium-dependent phosphate cotransporter; human; NAFPR; cancer; myopathy;

KW cell signalling disorder; phosphate regulation disorder; therapy; ss.
OS Homo sapiens.
PN MO9837198-A1.
PD 27-AUG-1998.
PF 24-FEB-1998; U03745.
PR 24-FEB-1997; US-805118.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Lal P.
DR WPI: 98-467561/40.

PT New Isolated human sodium-dependent phosphate co:transporter - used
PT to develop products for treating e.g. cancers, osteoporosis,
PT Alzheimer's disease, diabetes, encephalopathy, myopathy,
PT hypocalcaemia or hypoglycaemia.
PS Example V: Page 47; 66pp; English.

CC This sequence represents a fragment of the DNA encoding the human
CC sodium-dependent phosphate cotransporter (NAFPR) of the invention. NAFPR
CC and agonists of it can be used to treat or prevent disorders associated
CC with decreased phosphate levels, e.g. cancers of the kidney, disorders of
CC decreased phosphate levels including tumoural calcinosis, osteomalacia,
CC osteoporosis, familial hypophosphataemia, rickets, cystinuria,
CC nephrocalcinosis, glomerulonephritis, renal calculus, Alzheimer's
CC disease, diabetes mellitus, hereditary amyloidosis, myopathies including
CC progressive external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic
CC epilepsy, encephalopathy, cardiomyopathy, hypokalaemia, Goodpastures
CC syndrome, and disorders of cell signalling through cAMP, ATP, NADPH and
CC glucose-6-phosphate. Antagonists or inhibitors of NAFPR may be
CC administered to a subject to treat or prevent disorders associated with
CC increased phosphate levels, e.g. hypocalcaemia, hypocalcaemia, and
CC abnormal phosphate regulation in neurons, gastrointestinal tract and
CC liver. The products can also be used for detection, diagnosis and drug
CC screening. 272 BP; 67 A; 56 C; 58 G; 91 T;
SQ Sequence

alignment_scores:

Quality: 85.00 Length: 85
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x V43712 ..

Align seg 1/1 to: V43712 from: 1 to: 272

```
317 lLeTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMe 333
|||||
1 ATTAATATCAATGCTTAAGATATGCTCAAGGATATTCAGATTCTTCAT 50
|||||
333 tGlyAlaSerArgGlyPheSerSerIleAlaProValIleValProThrV 350
|||||
51 GGGAGCATCAAGAGAGATTTTCAGACATAGCACCTGATATGTATCCACTG 100
|||||
350 aLSerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnVal 365
|||||
101 TCAATGGATTTCTCTTAAGTCAGACCCCTGAGTTGGGTGGAGAAATGTC 150
|||||
367 PhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuI 383
|||||
151 TTTCTTTCGCGTTTGGCGTTAACTGTTAGACTACTCTTCACTCAT 200
|||||
383 ePheGlyGlyAlaAspValGlnGluTrpAlaLysGlyLysLeuThrA 400
|||||
201 ATTTGGAGAAAGCAGATGTCCAAGAAATGGGCTTAAGAGAAAACCTCATC 250
|||||
400 rGlu 401
|||||
251 GTTTA 255
|||||
```

seq_name: N_Geneseq_36:V57909

seq_documentation_block:

ID V57909 standard; cDNA: 2281 BP.
AC V57909;

CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. The present sequence
CC represents a hereditary haemochromatosis subregion from an hereditary
CC haemochromatosis (HH) affected individual. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BTF genes, which are homologues of the milk
CC protein butyrophilin (BT), and can be used in the production of agonists
CC and antagonists of BT function. Also described are: (1) a Rheset gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia.
SQ Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x V57903 ..

Align seg 1/1 to: V57903 from: 1 to: 237326

144 GlycylValCysCysLeuLeuTyrPhe 152

|||||
219417 GCGTGTCTGCTGCTGCTGCTGCTGCTC 219443

seq_name: N_Geneseq_36:V43714

seq_documentation_block:

ID V43714 standard; DNA; 25 BP.

AC V43714;

DT 30-NOV-1998 (first entry)

DE Primer for Sodium-dependent phosphate cotransporter coding sequence.

KW Sodium-dependent phosphate cotransporter; human; NAPT4; cancer; myopathy;

KW cell signalling disorder; phosphate regulation disorder; therapy;

KW PCR primer; ss.

OS Synthetic.

OS Homo sapiens.

PN MO9837198-A1.

PD 27-AUG-1998.

PF 24-FEB-1998; U03745.

PR 24-FEB-1997; US-805118.

PA (INC-) INCYTE PHARM INC.

PI Bandman O, Lal P;

DR WPI: 98-467561/40.

PT New isolated human sodium-dependent phosphate co:transporter - used

PT to develop products for treating e.g. cancers, osteoporosis,

PT Alzheimer's disease, diabetes, encephalopathy, myopathy,

PT hypocalcaemia or hypoglycaemia.

PS Example V: Page 36; 66pp; English.

CC This sequence represents a PCR primer for the DNA encoding the human

CC sodium-dependent phosphate cotransporter (NAPT4) of the invention. NAPT4

CC and agonists of it can be used to treat or prevent disorders associated

CC with decreased phosphate levels, e.g. cancers of the kidney, disorders of

CC osteoporosis, familial hypophosphataemia, rickets, cystinuria,

CC nephrocalcinosis, glomerulonephritis, renal calculus, Alzheimer's

CC disease, diabetes mellitus, hereditary amyloidosis, myopathies including

CC progressive external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic

CC epilepsy, encephalopathy, cardiomyopathy, hypocalcaemia, Goodpasture's

CC syndrome, and disorders of cell signalling through cAMP, ATP, NADPH and

CC glucose-6-phosphate. Antagonists or inhibitors of NAPT4 may be

CC administered to a subject to treat or prevent disorders associated with

CC Increased phosphate levels, e.g. hypocalcaemia, hypocalcaemia, and
CC abnormal phosphate regulation in neurons, gastrointestinal tract and
CC liver. The products can also be used for detection, diagnosis and drug
CC screening.
SQ Sequence 25 BP; 6 A; 6 C; 6 G; 7 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x V43714 ..

Align seg 1/1 to: V43714 from: 1 to: 25

338 GlyPheSerSerIleAlaProVal 345

|||||
2 GGATTTGACATGACACCTGTC 25

seq_name: N_Geneseq_36:Q14339

seq_documentation_block:

ID Q14339 standard; DNA; 1779 BP.

AC Q14339;

DT 17-JAN-1992 (first entry)

DE Human epithelin precursor.

KW ET; anti-sense RNA; growth regulation; inhibition; stimulation; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT misc_rna

FT /tag= a

FT /note= "claim 9, page 54"

PN WO9115510-A.

PD 17-OCT-1991.

PF 03-APR-1991; U02321.

PR 03-APR-1990; US-504508.

PR 13-MAR-1991; US-083796.

PA (BRIM) BRISTOL-MYERS SQUIB.

PI Shoyab M, Plozman GD;

DR WPI: 91-325168/44.

DR P-PSDB: R14326.

PT New cysteine-rich growth modulating proteins, epithelins - useful

PT as inhibitors of neoplastic cell growth and to promote wound

PT healing and treat psoriasis

PS Disclosure: Fig 22; 97pp; English.

CC ET-1 and ET-2 were isolated from rat kidneys and their amino acid

CC sequences determined. A full length rat ET cDNA (Q14338) was obtd.

CC by screening a rat kidney cDNA library in lambda gt10 with PCR

CC generated ET probes. These probes were also used to obtain the mouse

CC ET gene (Q14340) from a mouse T-cell genomic library.

CC ET DNA was also obtained from human sources (Q14339).

CC An anti-sense ribonucleic acid molecule complementary to the

CC indicated fragment in the features is also claimed.

CC ET-1 is a bifunctional growth regulator, capable of stimulating

CC the growth of some cell types while inhibiting the growth of others.

CC ET-2 is functionally similar to ET-1 w.r.t. growth inhibitory

CC bioactivity. In contrast, however, ET-2 is apparently not capable of

CC eliciting the growth stimulatory activity characteristic of ET-1 and,

CC in fact, antagonises this ET-1 activity.

CC See also Q14338-40, Q14952-53, R14328-9 and R15315-20.

CC Sequence 1779 BP; 304 A; 578 C; 547 G; 350 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x Q14339 ..

Align seg 1/1 to: Q14339 from: 1 to: 1779

281 ProSerSerAlaLeuIleValSer 288
 |||||
 368 CCATCCAGTCCCTGATGACTGACT 391

seq_name: N_Geneseq_36:V82825

seq_documentation_block:

ID V82825 standard; cDNA: 2095 BP.

AC V82825;

DT 15-MAR-1999 (first entry)

DE Human GP88 autocrine growth factor cDNA.

KW GP88; granuln; epithilin; human; growth factor; autocrine; tumour;

OS cancer; viral infection; antagonist; therapy; diagnosis; ss.

PI Key

FT CDS location/Qualifiers

FT 13..1794

PN W09852607-A1.

PD 26-NOV-1998.

PF 22-MAY-1998; U10555.

PR 16-DEC-1997; US-991862.

PR 23-MAY-1997; US-863079.

PA (SERR/) SERRERO G.

PI Serrero G;

DR WPI; 99-045276/04.

DR P-PSDB; W85475.

PT Composition containing antagonist of growth factor GP88 - useful for

PT treating cancer and viral diseases and also for diagnosing disease

PS Disclosure: Fig 9A; 86pp; English.

CC This human cDNA sequence includes a coding region for GP88 (see

CC W85474), an 88 kDa glycoprotein autocrine growth factor and

CC epithilin/granulin precursor that is expressed in a tightly

CC regulated manner in normal cells, is overexpressed and unregulated

CC in highly tumorigenic cells derived from normal cells, and which

CC acts as a stringently required growth stimulator for the

CC tumorigenic cells. Inhibition of GP88 expression or action in the

CC properties of the overproducing cells. Antagonists to GP88 are

CC used to treat diseases associated with increased expression of

CC GP88, particularly cancer but also viral infections. Fragments of

CC GP88 are used to raise specific antibodies (used as antagonists,

CC compounds to GP88-expressing cells) and to screen for antibodies.

CC Antisense oligonucleotides can also be used as antagonists.

CC Methods are provided for diagnosing disease, or determining

CC susceptibility to disease, resulting from altered GP88 activity.

Sequence 2095 BP; 364 A; 681 C; 625 G; 425 T;

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 8

Gaps: 0

US-09-391-958-1 x V82825 ..

Align seg 1/1 to: V82825 from: 1 to: 2095

281 ProSerSerAlaLeuIleValSer 288

|||||

380 CCATCCAGTCCCTGATGACTGACT 403

seq_name: N_Geneseq_36:Q56794

seq_documentation_block:

ID Q56794 standard; DNA: 2124 BP.

AC Q56794;

DT 22-APR-1994 (first entry)

DE Granulin coding sequence.

KW Granulin; keratinocytes; wound healing; inhibition; peptide;

KW granulocytes; leucocytes; ss.

OS Homo sapiens.

PI Key

FT misc_difference 1379..1381

FT /tag= a

FT /transl_except= ATG encodes Valine.

FT misc_difference 1640..1642

FT /tag= b

FT /transl_except= CAG encodes Glycine.

PN W09315195-A.

PD 05-AUG-1993.

PF 28-FEB-1992; CA0089.

PR 03-FEB-1992; US-829233.

PA (SOLO/) SOLOMON S.

PI Solomon S;

DR WPI; 93-320328/40.

DR P-PSDB; R48673.

PT New cysteine rich granulin peptide(s) from leucocyte(s) are

PT keratinocyte inhibitors useful topically for wound healing

PS Disclosure; Figure 4c; 53pp; English.

CC The granulin inhibits keratinocytes and is useful in formulations

CC for promoting the healing of wounds.

Sequence 2124 BP; 383 A; 685 C; 630 G; 426 T;

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 8

Gaps: 0

US-09-391-958-1 x Q56794 ..

Align seg 1/1 to: Q56794 from: 1 to: 2124

281 ProSerSerAlaLeuIleValSer 288

|||||

393 CCATCCAGTCCCTGATGACTGACT 416

seq_name: N_Geneseq_36:Q49052

seq_documentation_block:

ID Q49052 standard; DNA: 2198 BP.

AC Q49052;

DT 22-APR-1994 (first entry)

DE Granulin coding sequence.

KW Granulin; keratinocytes; wound healing; inhibition; peptide;

KW granulocytes; leucocytes; ss.

OS Homo sapiens.

PI Key

FT misc_difference 1453..1455

FT /tag= a

FT /transl_except= ATG encodes Valine.

FT misc_difference 1714..1716

FT /tag= a

FT /transl_except= CAG encodes Glycine.

PN W09315195-A.

PD 05-AUG-1993.

PF 28-FEB-1992; CA0089.

PR 03-FEB-1992; US-829233.

PA (SOLO/) SOLOMON S.

PI Solomon S;

DR WPI; 93-320328/40.

DR P-PSDB; R48673.

PT New cysteine rich granulin peptide(s) from leucocyte(s) are

PT keratinocyte inhibitors useful topically for wound healing

PS Disclosure; Figure 4c; 53pp; English.

CC The granulin inhibits keratinocytes and is useful in formulations

CC for promoting the healing of wounds.

Sequence 2198 BP; 398 A; 708 C; 646 G; 446 T;

CC resistance (e.g. to carbenazim and to ansamitocin) on transformed
CC microorganisms. See Q55405 and Q55406 for mutated sequences.
SQ Sequence 3445 BP; 723 A; 1061 C; 892 G; 769 T;

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x Q48230 ..

Align seg 1/1 to: Q48230 from: 1 to: 3445

278 GlySerLeuProSerSerAlaLeu 285
|||||
2360 GGTTCCCTCCCTCGTCGACCTT 2383

seq_name: N_Geneseq_36:Q55405

seq_documentation_block:

ID Q55405 standard; DNA; 3445 BP.

AC Q55405;
DE 22-FEB-1994 (first entry)
DE A.chrysogenum beta-tubulin Ile(100) mutant coding sequence.
KW Beta tubulin; mutant; chemical resistance; selective marker;
KW cephalosporin; antibiotic production; ds.
OS Acromonium chrysogenum.
FH Key
FT exon
FT 1. .1298
FT Location/Qualifiers
FT /*tag= a
FT /note= "ATG initiation codon is located at
FT nucleotides 1287. .1289"
FT 1299. .1460
FT /*tag= b
FT /number= 1
FT 1461. .1484
FT /*tag= c
FT /number= 2
FT 1485. .1551
FT /*tag= d
FT /number= 2
FT 1552. .1674
FT /*tag= e
FT /number= 3
FT 1675. .1748
FT /*tag= f
FT /number= 3
FT 1749. .2539
FT /*tag= g
FT /number= 4
FT /note= "wild-type AAC (Asn) codon at position
FT 1887. .1889 is mutated to ATC (Ile) codon"
FT 2540. .2602
FT /*tag= h
FT /number= 4
FT 2603. .3445
FT /*tag= i
FT /number= 5
FT /note= "TAA termination codon is located at
FT nucleotides 2994. .2996"

PN J05192157-A.

PD 03-AUG-1993.

PF 26-MAY-1992; 133384.

PR 27-MAY-1991; JP-121276.

PA (TAKE) TAKEDA CHEM IND LTD.

DR WPI: 93-277472/35.

DR P-PSDB: R48200.

PT DNA fragment contg. DNA coding mutant beta-tubulin - originates
PT from Acromonium chrysogenum, used as selective marker for
PT transformation of A.chrysogenum

PS Claim 3 and Example 6; Fig 4-6 and Fig 7; 16pp; Japanese.
CC The wild-type coding sequence for beta-tubulin was isolated from
CC Acromonium chrysogenum ATCC 11550 and sequenced (Q48230). Primers
CC CTU-3 and CTU-6 were used to introduce mutations at codon 100 (Asn
CC to Ile) and 167 (Phe to Tyr), respectively. Expression of the
CC mutant proteins encoded by these sequences confers chemical
CC resistance (e.g. to carbenazim and to ansamitocin) on transformed
CC microorganisms. See also Q55406.
SQ Sequence 3445 BP; 722 A; 1061 C; 892 G; 770 T;

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x Q55405 ..

Align seg 1/1 to: Q55405 from: 1 to: 3445

278 GlySerLeuProSerSerAlaLeu 285
|||||
2360 GGTTCCCTCCCTCGTCGACCTT 2383

seq_name: N_Geneseq_36:Q55406

seq_documentation_block:

ID Q55406 standard; DNA; 3445 BP.

AC Q55406;
DE 22-FEB-1994 (first entry)
DE A.chrysogenum beta-tubulin Tyr(167) mutant coding sequence.
KW Beta tubulin; mutant; chemical resistance; selective marker;
KW cephalosporin; antibiotic production; ds.
OS Acromonium chrysogenum.
FH Key
FT exon
FT 1. .1298
FT Location/Qualifiers
FT /*tag= a
FT /note= "ATG initiation codon is located at
FT nucleotides 1287. .1289"
FT 1299. .1460
FT /*tag= b
FT /number= 1
FT 1461. .1484
FT /*tag= c
FT /number= 2
FT 1485. .1551
FT /*tag= d
FT /number= 2
FT 1552. .1674
FT /*tag= e
FT /number= 3
FT 1675. .1748
FT /*tag= f
FT /number= 3
FT 1749. .2539
FT /*tag= g
FT /number= 4
FT /note= "wild-type TTC (Phe) codon at position
FT 2088. .2090 is mutated to TAC (Tyr) codon"
FT 2540. .2602
FT /*tag= h
FT /number= 4
FT 2603. .3445
FT /*tag= i
FT /number= 5
FT /note= "TAA termination codon is located at
FT nucleotides 2994. .2996"

PN J05192157-A.

PD 03-AUG-1993.

PF 26-MAY-1992; 133384.

PR 27-MAY-1991; JP-121276.

PA (TAKE) TAKEDA CHEM IND LTD.
 DR WPI: 93-277472/35.
 DR P-PSDB: R48201.
 PT DNA fragment contg. DNA coding mutant beta-tubulin - originates
 PT from Acremonium chrysogenum, used as selective marker for
 PT transformation of A. chrysogenum
 PS Claim 3 and Example 6: Fig 4-6 and Fig 7; 16pp; Japanese.
 CC The wild-type coding sequence for beta-tubulin was isolated from
 CC Acremonium chrysogenum ATCC 11550 and sequenced (Q48230). Primers
 CC CTU-3 and CTU-6 were used to introduce mutations at codon 100 (Asn
 CC to Ile) and 167 (Phe to Tyr), respectively. Expression of the
 CC mutant proteins encoded by these sequences confers chemical
 CC resistance (e.g. to carbendazim and to ansamitocin) on transformed
 CC microorganisms. See also Q55405.
 SQ Sequence 3445 BP; 724 A; 1061 C; 892 G; 768 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-391-958-1 x Q55406 ..

Align seg 1/1 to: Q55406 from: 1 to: 3445

278 GlySerLeuProSerSerAlaLeu 285
 ||||||||||||||||||||
 2360 GGTTCCCTTCCCTGCTGCACCTT 2383

THIS PAGE BLANK (USPTO)

OM of: US-09-391-958-1 to: Issued_Patents_NA: * out_format : pfs

About: Results were produced by the Gencore software, version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

```

Command line parameters:
-MODEL=frame+ p2n.model
-O=/cgn2.1/USPPO.spool/USO09319158/runat.24052000_213606_8858/app-query.fasta.1.1
-DEV=xlip
-DB=Issued_Patents_NA_QPMT-fastad-SUFFIX=oligo.tni
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPC=0.000
-LOOPEXT=0.000 -FGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000
-XGAPEXT=60.000 -FGAPEXT=7.000 -XGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCLIST=200
-THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFWT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=100000 -USER=USO09319158 -NCPD=6
-ICPD=3 -NO_XLIPX -WAIT -THREADS=1

```

```
Search Information block:
Query: US-09-391-958-1
Query length: 401
Database: Issued_Patents_NA:
Database sequences: 230463
Database length: 64992525
Search time (sec): 47.270000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_1st:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	..	1643	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-805-118-2 +			401.00	7577.11	0.0		1780	I
/cgn2.6/prodata/1/lna/5C_COMB.seq:US-08-824-394A-19 +			241.00	4539.51	2.2e-245		1561	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-805-118-5 +			85.00	1593.39	2.8e-81		212	I
/cgn2.6/prodata/1/lna/5C_COMB.seq:US-08-724-394A-18 +			9.00	134.00	5.3e		226	I
/cgn2.6/prodata/1/lna/5C_COMB.seq:US-08-724-394A-20 +			9.00	96.78	634.49		26540	I
/cgn2.6/prodata/1/lna/5C_COMB.seq:US-08-724-394A-21 +			9.00	96.78	634.49		246540	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-724-394A-22 -			9.00	96.78	634.49		246540	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-805-118-7 +			8.00	150.80	0.6222		25	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-07-668-648-3 +			8.00	116.94	47.83		1779	I
/cgn2.6/prodata/1/lna/5C_COMB.seq:US-08-829-939-3 +			8.00	116.94	47.83		1779	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-431-333-3 +			8.00	116.94	47.83		1779	I
/cgn2.6/prodata/1/lna/POTUS_COMB.seq:PCRT-US01-02321-3 +			8.00	116.94	47.83		1779	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-805-118-6 +			7.00	131.81	7.10		25	I
/cgn2.6/prodata/1/lna/POTUS_COMB.seq:PCRT-US94-01149-51 -			7.00	127.89	11.75		41	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-825-243A-4 +			7.00	126.31	14.38		50	I
/cgn2.6/prodata/1/lna/6_COMB.seq:US-08-433-522A-50 -			7.00	121.74	25.86		89	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-825-243A-2 +			7.00	113.26	76.72		259	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-406-248-5 +			7.00	112.17	88.20		257	I
/cgn2.6/prodata/1/lna/POTUS_COMB.seq:PCRT-US94-05150-24 +			7.00	111.00	102.43		344	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-474-542A-102 -			7.00	110.38	110.92		372	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-474-542A-201 +			7.00	108.68	137.99		461	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-457-648-291 +			7.00	108.68	137.99		461	I
/cgn2.6/prodata/1/lna/backfile1.seq:52252348-7 -			7.00	107.21	166.66		555	I
/cgn2.6/prodata/1/lna/backfile1.seq:52252348-8 -			7.00	105.82	175.29		583	I
/cgn2.6/prodata/1/lna/backfile1.seq:52252348-9 -			7.00	106.60	180.15		599	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-816-233C-1 +			7.00	104.54	234.49		776	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-117-083-7 +			7.00	104.40	238.79		790	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-440-84E-1 +			7.00	102.73	285.84		975	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-07-965-274-3 +			7.00	102.53	303.56		1000	I
/cgn2.6/prodata/1/lna/POTUS_COMB.seq:PCRT-US95-05616-3 +			7.00	101.70	337.59		1110	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-020-245A-3 -			7.00	101.10	364.86		1198	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-396-957A-3 -			7.00	101.10	364.86		1198	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-08-265-423A-3 -			7.00	100.64	386.89		1269	I
/cgn2.6/prodata/1/lna/POTUS_COMB.seq:PCRT-US95-09065-4 +			7.00	100.64	386.89		1269	I
/cgn2.6/prodata/1/lna/POTUS_COMB.seq:PCRT-US93-05640-10 +			7.00	99.51	447.19		14661	I
/cgn2.6/prodata/1/lna/5A_COMB.seq:US-07-575-536B-1 -			7.00	99.00	477.70		1561	I
/cgn2.6/prodata/1/lna/backfile1.seq:52252348-4 -			7.00	99.00	477.70		1561	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-922-170B-9 -			7.00	98.22	527.60		1721	I
/cgn2.6/prodata/1/lna/5D_COMB.seq:US-08-922-170B-11 -			7.00	98.22	527.60		1721	I

seq_name	seq	US	-08	-805	-118	-2
/cgn2_6/prodata/1/1na/5D_COMB	seq	US	-08	-537	-899	-2
/cgn2_6/prodata/1/1na/5C_COMB	seq	US	-08	-653	-899	-2
/cgn2_6/prodata/1/1na/5C_COMB	seq	US	-08	-558	-970	-1
/cgn2_6/prodata/1/1na/5C_COMB	seq	US	-08	-458	-970	-1
/cgn2_6/prodata/1/1na/5C_COMB	seq	US	-08	-105	-230	-8
/cgn2_6/prodata/1/1na/5C_COMB	seq	US	-08	-110	-990	-8
/cgn2_6/prodata/1/1na/5D_COMB	seq	US	-08	-902	-853	-4
/cgn2_6/prodata/1/1na/5D_COMB	seq	US	-08	-902	-853	-4
/cgn2_6/prodata/1/1na/5D_COMB	seq	US	-08	-805	-118	-2

```

seq_documentation_block:
: Sequence 2, Application US/08805118
: Patent No. 5985604
:
GENERAL INFORMATION:
: APPLICANT: Lal, Preeti
: APPLICANT: Bandman, Olga
: TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
: TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
: NUMBER OF SEQUENCES: 7
:
CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: US
: ZIP: 94304
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/805,118
: FILING DATE: Filed Herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0221 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1643 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: BRAIT0702
: CLONE: 754412
:
US-08-805-118--2
:
alignment_scores:
: Quality: 401.00 Length: 401
: Ratio: 1.000 Gaps: 0
: Percent Similarity: 100.000 Percent Identity: 100.000
:
alignment_block:
US-09-391-958-1 x US-08-805-118-2 ..
Align seg 1/1 to: US-08-805-118--2 from: 1 to: 1643
:
1 MetGlnValAspGluThrLeuIleProArgLysValProSerLeuCysSe 17
|||||
235 ATGCAAGTGGATGAGACACTGATCCCAAGAAAGTTCACAAATTGATTC 284
|||||
17 TAlAaGtGtYrGtYlLeAlaLeuValLeuHIsPhEcYsAsnPhetHrThI 34
|||||
285 TGCTGCGATGGAATAGCCCTGCTGTACATTTCGCAATTTCACAAACCA 334
|||||

```

34 lealGlnAsnValIleMetAsnIleThrMetValAlaMetValAsnSer 50
 335 TAGCAAAAGTGCATCATGAAACATCACCATGGTGCATGCAACAGC 384
 51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
 385 ACMAAGCCCTCAATCCAGCTCATATGATTCCTGTGAGGTGCTGCCCTTGA 434
 67 pSerPheGlyLeuSerLysAlaProLysSerLeuProAlaLysSerS 84
 435 CTCATTTGGTGGCCTAAGTAAGCCCAAGATCTTCTTCCAAAGTCT 484
 84 erlleuGlyGlyGlnPheAlaIleTrpGluArgTrpGlyProProGln 100
 485 CAATACCTGGGGGTGAGTTGCAATTGGGAAGGTGGGGCCCTCCACAA 534
 101 GluArgSerArgLeuGlySerLysAlaLeuSerLysMetLeuGlyCys 117
 535 GAACGAAGCAGACTCTGCAGCATGCTTATCAGGAATGTTACTGGGATG 584
 117 sPheThrAlaIleLeuIleGlyPheIleSerGluThrLeuGlyTrp 134
 585 CTTTACTGGCCATCTCATAGGTGGCTTCATTAGGAACCTTGGGTGGC 634
 134 ropPheValPheTrpIlePheGlyGlyValGlyCysValCysLeuLeu 150
 635 CTTTGTCTCTATATCTTGGAGGTGTGGCTGTCTGCTGCTCTCTC 684
 151 TrpPheValValIleTrpAspAspProValSerTrpProTrpIleSerTh 167
 685 TGGTTTGTGATTTATGATGAGCCCGTTTCCATTCATGATGATTAAGC 734
 167 rSerGluGlyGluTrpIleIleSerSerLeuLysGlnGlnValGlySers 184
 735 CTCGAAAAAAGATATCATATCTCTTGAACAAACAGGTGGGCTCT 784
 184 erlyGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTrp 200
 785 CTAAAGCAGCTCTCCCATTAAGCTATGCTCAGATCTCTAACCATTTGG 834
 201 SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMetVa 217
 835 TCCATATGTTTAGGCTGTTTACAGCATCATAGTTAGTTAGCAGCATGGT 884
 217 lValTrpIleProThrTrpIleSerSerValTrpHisValAsnIleArgA 234
 885 TGTATACATACCAACTTACATCAGCTGTGTACCATGTTAACATCAGAG 934
 234 spAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
 935 ACAATGGACTGTATCTGCCCTTCTTATTTGTTGCTGGGTCAATAGGC 984
 251 MetValGlyGlyTrpLeuAlaAspPheLeuThrLysLysPheArgLe 267
 985 ATGGTGGGAGGCTATCTGGCAGATTCTTCTTAACCAAAAAGTTTAGACT 1034
 267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSera 284
 1035 CATCACTGTGAGGAAATTTGCCACAAATTTAGGAAGTCTCCCTCTTCAG 1084
 284 lAlaLeuIleValSerLeuProTrpLysAsnSerGlyTrpIleThrAlaThr 300
 1085 CACTCATTTGTCTGTGCTTACCTCAATTCGGGCTATATTCACAGCACT 1134
 301 AlAlaLeuThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGlyTrp 317
 1135 GCCTTGTGAGCGCTCTTTCGCGATTAAAGCACAATGTGTCAGATCAGGAT 1184
 317 eTrpIleAsnValLeuAspIleAlaProArgTrpSerSerPheLeuMetG 334
 1185 TTATATCATATGCTTAGATATTGCTCCAAAGGATTTCCAGATTTTTCATGG 1234

334 lYAlaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
 1235 GAGCATCAAGAGGATTTTCGAGCATAGCAGCTGATTCATGTGCCAGTGTCTC 1284
 351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValAlp 367
 1285 AGCGGATTTCTTCTTACTGACGAGACCTGAGTTGGTGGAGGAATGTCTT 1334
 367 ePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTrpLeuIleP 384
 1335 CTTCTCTCTGTTTSCGCTTAACCTGTTAGAGCATACCTTCTTACTCATAT 1384
 384 heGlyGluAlaAspValGlnGluTrpAlaLysGluArgLysLeuThrArg 400
 1385 TTGAGAGAGCAGATGTCAGAAATGGCTTAAGAGAGAGAAAACTACACTGT 1434
 401 leu 401
 1435 TTA 1437
 seq_name: /cgn2.6/plodata/1/lna/5C_COMB.seq:us-08-724-394A-19
 seq_documentation_block:
 ; Sequence 19, Application US/08724394A
 ; Patent No. 5872237
 ; GENERAL INFORMATION:
 ; APPLICANT: Feder, John N.
 ; APPLICANT: Krommal, Gregory S.
 ; APPLICANT: Lauer, Peter M.
 ; APPLICANT: Ruddy, David A.
 ; APPLICANT: Thomas, Winston
 ; APPLICANT: Tsuchihashi, Zenta
 ; APPLICANT: Wolff, Roger K.
 ; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
 ; TITLE OF INVENTION: Sequences and Antibodies Thereo
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/724,394A
 ; FILING DATE: 01-OCT-1996
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fitts, Renee A.
 ; REGISTRATION NUMBER: 35,136
 ; REFERENCE/DOCKET NUMBER: 017957-000100
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-576-0200
 ; TELEFAX: 415-576-0300
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1780 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..1780
 ; OTHER INFORMATION: /note="cDNA 22E"
 US-08-724-394A-19

alignment_scores: Quality: 241.00 Length: 241
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x US-08-724-394A-19 ..

Align seg 1/1 to: US-08-724-394A-19 from: 1 to: 1780

```
161 SerTyrProTrpIleSerThrSerGluLysGluTyrIleIleSerSerLe 177
|||||
857 TCCATTCATGAGTAAAGCACCCTCAGAAAAAGAAATACATCATATCTCTCT 906
|||||
177 uLysGlnGlnValGlySerSerLysGlnProLeuProIleLysAlaMetL 194
|||||
907 GAAACAACAGGTGGGCTCTTCTTAAAGCAGCCCTTCCATCAAAAGCTATGC 956
|||||
194 euArGSerLeuProIleTrpSerIleCysLeuGlyCysPheSerHisGln 210
|||||
957 TCAGATCTCAACCATTTGGCTCATGTATTAGCGCTGTTCCAGCCATCAA 1006
|||||
211 TrpLeuValSerThrMetValValTyrIleProThrTyrIleSerSerVa 227
|||||
1007 TGGTTAGTAGCAACAATGGTTGTATACATACCACTTACATCAGCTCTGT 1056
|||||
227 lTyHisValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuProPheI 244
|||||
1057 GTACCAATGTTAAACATCAAGACAAATGAGACTTCTATGCTCCCTCTTTA 1106
|||||
244 lValAlaTrpValIleGlyMetValGlyGlyTyrLeuAlaAspPheLeu 260
|||||
1107 TTGTGGCTGGGTATAGCATGGTGGAGAGCTATCTGGCGATTTCCCTT 1156
|||||
261 LeuThrLysLysPheArgLeuIleThrValArgLysIleAlaThrIleLe 277
|||||
1157 CTAACCAAAAAAGTTTACATCATCATCATGTGAGAAAAATGCCACAATT 1206
|||||
277 uGlySerLeuProSerSerAlaLeuIleValSerLeuProTyrLeuAsn 294
|||||
1207 AGGAAGTCTCCCTCTTACAGACATCATGCTGCTGCTTACCTCAAT 1256
|||||
294 eArgLysIleThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSer 310
|||||
1257 CCGGCTATATCAGCAAGCACTGCTGCTGAGCGCTCTCTGGGGAATTAA 1306
|||||
311 ThrLeuCysGlnSerGlyIleTyrIleAsnValLeuAspIleAlaProAr 327
|||||
1307 ACATTTGTGTGATGTCAGGATTTATATCAATGCTTAGATATTGCTCCA 1356
|||||
327 gTyrSerSerPheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaP 344
|||||
1357 GTATTTCAGATTTTCTCATGGAGCATCAAGAGATTTTCGAGCATAGAC 1406
|||||
344 roValIleValProThrValSerGlyPheLeuLeuSerGlnAspProGlu 360
|||||
1407 CTGTGATATTGATCCCATGCTCAGCGGATTTCTTCTTAGTCAGAGACCT 1456
|||||
361 PheGlyTrpArgAsnValPhePheLeuLeuPheAlaValAsnLeuLeuG 377
|||||
1457 TTTGGGTGAGAGATGCTTCTTCTGCTGTTGCGGTTAACTGTGTAGG 1506
|||||
377 yLeuLeuPheTyrLeuIlePheGlyGlyAlaAspValGlnGluTrpAla 394
|||||
1507 ACTACTCTTCTTACCTCATATTGGAAGAGCATGTCACAAAGATGGGCTA 1556
|||||
394 ySgIuArgLysLeuThrArgLeu 401
|||||
1557 AAGAGAGAAAACTCATCTGTTTA 1579
```

seq_name: /cgn2_6/ptodata/1/lna/5D.COMB.seq: US-08-805-118-5
seq_documentation_block:

```
? Sequence 5, Application US/08805118
? Patent No. 5985604
? GENERAL INFORMATION:
? APPLICANT: Lal, Preeti
? APPLICANT: Bandman, Olga
? TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
? TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
? ADDRESS: Incyte Pharmaceuticals, Inc.
? STREET: 3174 Porter Drive
? CITY: Palo Alto
? STATE: CA
? COUNTRY: US
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSRO Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/805,118
? FILING DATE: Filed Herewith
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J.
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0221 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-4166
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 272 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: BRAINFUT02
? CLONE: 754412
? US-08-805-118-5
```

alignment_scores: Quality: 85.00 Length: 85
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x US-08-805-118-5 ..

Align seg 1/1 to: US-08-805-118-5 from: 1 to: 272

```
317 lIeTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeu 333
|||||
1 ATTATATATCAATGCTTATGATATTGCTTCCAAAGATTCAGATTCTCTAT 50
|||||
333 tGAlaSerArgGlyPheSerSerIleAlaProValIleValProThrV 350
|||||
51 GGGAGCATCAAGAGGATTTTCAGAGCATGCACTGCTCATTTGATCCACTG 100
|||||
101 TCAGTGGATTTCTCTTCTTAGTCAGGACCTGAGATTGGGTGAGAGATGTC 150
|||||
350 aISerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnVal 366
|||||
367 PhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuI 383
|||||
151 TTCTTCTTGGCTTTGGCCGTTAACTGTTAGGACTACTCTTCTTACTCAT 200
|||||
383 ePheGlyGlnAlaAspValGlnGluTrpAlaLysGlnArgLysLeuThr 400
```

|||||
201 ATTGAGAGAGATGTCAGATGGCTTAAGAGAAACTACTC 250
400 TGTAA 401
|||||
251 GTTTA 255

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:us-08-724-394A-18

seq_documentation_block:

; Sequence 18, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

; APPLICANT: Kronmal, Gregory S.

; APPLICANT: Laue, Peter M.

; APPLICANT: Ruddy, David A.

; APPLICANT: Thomas, Winston

; APPLICANT: Tsuchihashi, Zenta

; APPLICANT: Wolff, Roger K.

; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/724.394A

; FILING DATE: 01-OCT-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Fitts, Renee A.

; REGISTRATION NUMBER: 35,136

; REFERENCE/DOCKET NUMBER: 017957-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2266 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1..2266

; OTHER INFORMATION: /note="cDNA 22B"

; US-08-724-394A-18

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:us-08-724-394A-20

seq_documentation_block:

; Sequence 20, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

; APPLICANT: Kronmal, Gregory S.

; APPLICANT: Laue, Peter M.

; APPLICANT: Ruddy, David A.

; APPLICANT: Thomas, Winston

; APPLICANT: Tsuchihashi, Zenta

; APPLICANT: Wolff, Roger K.

; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/724.394A

; FILING DATE: 01-OCT-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Fitts, Renee A.

; REGISTRATION NUMBER: 35,136

; REFERENCE/DOCKET NUMBER: 017957-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 246240 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1..246240

; OTHER INFORMATION: /note="HLA-H.CONTIG"

; US-08-724-394A-20

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x US-08-724-394A-20/rev ..

Align seg 1/1 to reverse of: US-08-724-394A-20 from: 1 to: 246240

144 GlycysValcysCysLeuLeuTrrpPhe 152

|||||

22303 GGCTGTCTGCTGCTCTCCATGATGTTTC 22277

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:us-08-724-394A-21

seq_documentation_block:

; Sequence 21, Application US/08724394A

; Patent No. 5872237

```

: GENERAL INFORMATION:
: APPLICANT: Feder, John N.
: APPLICANT: Krommal, Gregory S.
: APPLICANT: Lauer, Peter M.
: APPLICANT: Ruddy, David A.
: APPLICANT: Thomas, Winston
: APPLICANT: Tsuchihashi, Zenta
: APPLICANT: Wolff, Roger K.
: TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
: TITLE OF INVENTION: Sequences and Antibodies Thereto
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
: STREET: Two Embarcadero Center, 8th Floor
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/724,394A
: FILING DATE: 01-OCT-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Fitts, Renee A.
: REGISTRATION NUMBER: 35,136
: REFERENCE/DOCKET NUMBER: 017957-000100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-576-0200
: TELEFAX: 415-576-0300
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 246240 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1..246240
: OTHER INFORMATION: /note="HLA-H.CONTIG"
: US-08-724-394A-21

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-391-958-1 x US-08-724-394A-21/rev ..

Align seg 1/1 to reverse of: US-08-724-394A-21 from: 1 to: 246240

144 G1cYsValcYsCysLeuLeuTrpPhe 152
|||||
22303 GGCTGTCGTCTGCTGTCCTATGTC 22277

seq_name: /cgn2.6/ptodata/1/lna/5C.COMB.seq:US-08-724-394A-22

seq_documentation_block:
: Sequence 22, Application US/08724394A
: Patent No. 5872237
: GENERAL INFORMATION:
: APPLICANT: Feder, John N.
: APPLICANT: Krommal, Gregory S.
: APPLICANT: Lauer, Peter M.
: APPLICANT: Ruddy, David A.
: APPLICANT: Thomas, Winston
```

```

: APPLICANT: Tsuchihashi, Zenta
: APPLICANT: Wolff, Roger K.
: TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
: TITLE OF INVENTION: Sequences and Antibodies Thereto
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
: STREET: Two Embarcadero Center, 8th Floor
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/724,394A
: FILING DATE: 01-OCT-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Fitts, Renee A.
: REGISTRATION NUMBER: 35,136
: REFERENCE/DOCKET NUMBER: 017957-000100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-576-0200
: TELEFAX: 415-576-0300
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 246240 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1..246240
: OTHER INFORMATION: /note="HLA-H.CONTIG"
: US-08-724-394A-22

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-391-958-1 x US-08-724-394A-22/rev ..

Align seg 1/1 to reverse of: US-08-724-394A-22 from: 1 to: 246240

144 G1cYsValcYsCysLeuLeuTrpPhe 152
|||||
22303 GGCTGTCGTCTGCTGTCCTATGTC 22277

seq_name: /cgn2.6/ptodata/1/lna/5D.COMB.seq:US-08-805-118-7

seq_documentation_block:
: Sequence 7, Application US/08805118
: Patent No. 5985604
: GENERAL INFORMATION:
: APPLICANT: Lal, Preeti
: APPLICANT: Bandman, Olga
: TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
: TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: US
```

ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,118
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0221 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: XLF
US-08-805-118-7

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x US-08-805-118-7 ..

Align seg 1/1 to: US-08-805-118-7 from: 1 to: 25

338 GlyPheSerIleAlaProVal 345
|||||
2 GGATTTTCGACGATAGCAGCTGTC 25

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-668-648-3

seq_documentation_block:

Sequence 3, Application US/07668648
Patent No. 5416192
GENERAL INFORMATION:
APPLICANT: Shoyab, Mohammed
APPLICANT: Plozman, Gregory D.
TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH
TITLE OF INVENTION: MODULATING PROTEINS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
City: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/668,648
FILING DATE: 19910819
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-161-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212) 869-9741
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1779 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Kidney
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1779
US-07-668-648-3

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x US-07-668-648-3 ..

Align seg 1/1 to: US-07-668-648-3 from: 1 to: 1779

281 ProSerAlaIleValSer 288
|||||
368 CCATCCAGTGCCTGATGTCAGT 391

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-429-998-3

seq_documentation_block:

Sequence 3, Application US/08429998
Patent No. 5885961
GENERAL INFORMATION:
APPLICANT: Shoyab, Mohammed
APPLICANT: Plozman, Gregory D.
TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH
TITLE OF INVENTION: MODULATING PROTEINS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
City: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/429,998
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/668,648
FILING DATE: 13-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-161-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212) 869-9741

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1779 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: kidney
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1779
; US-08-429-998-3

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-391-958-1 x US-08-429-998-3  ..

Align seg 1/1 to: US-08-429-998-3 from: 1 to: 1779

281 ProserSerAlaLeuIleValSer 288
|||||
368 CCATCCAGTCCCTGATAGTCAGT 391

seq_name: /cgn2_6/prodata/1/lna/5D_COMB.seq:US-08-431-333-3

seq_documentation_block:
; Sequence 3, Application US/08431333
; Patent No. 5965723
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Plozman, Gregory D.
; TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH
; TITLE OF INVENTION: MODULATING PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/431,333
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/668,648
; FILING DATE: 13-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mastro, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-161-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1779 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: kidney
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1779
; US-08-431-333-3

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-391-958-1 x US-08-431-333-3  ..

Align seg 1/1 to: US-08-431-333-3 from: 1 to: 1779

281 ProserSerAlaLeuIleValSer 288
|||||
368 CCATCCAGTCCCTGATAGTCAGT 391

seq_name: /cgn2_6/prodata/1/lna/PCTUS_COMB.seq:PCT-US91-02321-3

seq_documentation_block:
; Sequence 3, Application PC/US9102321
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Plozman, Gregory D.
; TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH
; TITLE OF INVENTION: MODULATING PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Bristol-Myers Squibb Company
; STREET: 3005 First Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/02321
; FILING DATE: 19910403
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Poor, Brian W.
; REGISTRATION NUMBER: 32,928
; REFERENCE/DOCKET NUMBER: ON0071A-PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)728-4800
; TELEFAX: (206)448-4775
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1779 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: kidney
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1779
; PCT-US91-02321-3
```

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x PCT-US91-02321-3 ..

Align seg 1/1 to: PCT-US91-02321-3 from: 1 to: 1779

281 ProSerSerAlaLeuIleValSer 288
|||||
368 CCATCCAGTCCCTGATGATGTCAGT 391

seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-805-118-6

seq_documentation_block:

```

; Sequence 6, Application US/08805118
; Patent No. 5985604
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
; TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,118
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0221 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: XLR
; US-08-805-118-6

```

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x US-08-805-118-6/rev ..

Align seg 1/1 to reverse of: US-08-805-118-6 from: 1 to: 25

330 SerPheLeuMetGlyAlaSer 336

|||||
23 AGTTTCATGGAGCATCA 3

seq_name: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:PCT-US94-01149-51

seq_documentation_block:

```

; Sequence 51, Application PC/TUS9401149
; GENERAL INFORMATION:
; APPLICANT: Shatzman, Allan
; APPLICANT: Scott, Miller
; APPLICANT: Dillon, Susan B.
; APPLICANT: Kane, James
; TITLE OF INVENTION: Vaccinal Polypeptides
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation - Corporate
; ADDRESSEE: Patents
; STREET: U. S. Mailcode UW2220 - 709 Swedeland Road
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 149,150
; FILING DATE: 05-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 013,415
; FILING DATE: 01-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 108,914
; FILING DATE: 18-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 837,773
; FILING DATE: 18-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 751,896
; FILING DATE: 30-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 387,200
; FILING DATE: 28-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 238,801
; FILING DATE: 02-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 645,732
; FILING DATE: 30-AUG-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50134 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-270-5096
; TELEFAX: 215-270-5090
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; PCT-US94-01149-51

```

alignment_scores:

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:50:46 ; Search time 51.62 Seconds
(without alignments)
455.454 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 401
Sequence: 1 MOVDETLIPRKVPSLCSARY.....LIFGEADVQEMAKERKLTRL 401

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 168808 seqs, 58629743 residues

Word size : 0

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : PIR_63:*

1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.0	333	2 G72002	aspartate dehydrog
2	8	2.0	376	2 D72647	hypothetical prote
3	8	2.0	836	2 T08424	tweety protein - f
4	7	1.7	67	2 C75397	hypothetical prote
5	7	1.7	68	2 S35174	cytochrome P450 (c
6	7	1.7	97	2 T15891	hypothetical prote
7	7	1.7	98	1 W7MLHS	E7 protein - human
8	7	1.7	128	2 E71061	hypothetical prote
9	7	1.7	142	2 T08514	outer membrane pro
10	7	1.7	151	2 H64325	hypothetical prote
11	7	1.7	165	2 A45192	ribosomal protein
12	7	1.7	172	2 S75094	hypothetical prote
13	7	1.7	212	2 S56410	hypothetical prote
14	7	1.7	220	2 S43291	FLY3/FLK2 ligand (
15	7	1.7	220	2 I58343	fil3 ligand isofor
16	7	1.7	235	2 A72359	conserved hypothet
17	7	1.7	245	2 G75097	hypothetical prote
18	7	1.7	249	2 I50464	MHC class II beta
19	7	1.7	266	1 CDWT	chlorophyll a/b-di
20	7	1.7	273	2 A41607	ubiquinol--cytochr
21	7	1.7	304	2 B60444	Na+/K+-exchanging
22	7	1.7	305	1 S52161	probable fructokin
23	7	1.7	310	2 A70421	fructose-1,6-bisph
24	7	1.7	324	2 DA2951	alkanal monooxygen
25	7	1.7	324	2 D38448	alkanal monooxygen
26	7	1.7	327	2 C35411	alkanal monooxygen
27	7	1.7	327	2 C37898	alkanal monooxygen
28	7	1.7	343	2 T33989	hypothetical prote
29	7	1.7	344	2 JC5942	chemokine receptor

30	7	1.7	379	1 S04040	alcohol dehydrogen
31	7	1.7	385	2 T31493	hypothetical prote
32	7	1.7	400	2 A39254	inositol-1,4-bisph
33	7	1.7	429	2 T09888	hypothetical prote
34	7	1.7	435	2 E70586	hypothetical prote
35	7	1.7	444	2 A75319	phosphoglucosylase
36	7	1.7	465	2 I39473	Na+-dependent phos
37	7	1.7	465	2 A56410	sodium/phosphate t
38	7	1.7	465	2 S68915	sodium/phosphate t
39	7	1.7	467	2 A48916	sodium phosphate t
40	7	1.7	484	4 A32761	hypothetical Alzhe
41	7	1.7	492	2 B70617	hypothetical prote
42	7	1.7	512	1 A70201	virulence factor m
43	7	1.7	512	2 T38422	probable chromatin
44	7	1.7	524	2 A75588	probable protein k
45	7	1.7	537	2 T33955	hypothetical prote

ALIGNMENTS

RESULT 1
G72002
aspartate dehydrogenase - Chlamydia pneumoniae (strain CML029)
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: G72002
R:Kaltman, S.; Mitchell, W.; Marathe, R.; Iammell, C.; Fan, J.; Olinger, L.; Grimwood, A.; Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: G72002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <ARN>
A:Cross-references: GB:AE001685; GB:AE001363; NID:g4377378; PIDN:AAD19185.1; PID:g437
A:Experimental source: strain CML029
C:Genetics:
A:Gene: asd
C:Superfamily: yeast aspartate-semialdehyde dehydrogenase

Query Match 2.0%; Score 8; DB 2; Length 333;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 GSSKQPLP 189
DB 206 GSSKQPLP 213

RESULT 2
D72647
hypothetical protein APE0610 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72647
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A:Reference number: A72450; MUID:99310339
A:Accession: D72647
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <KAW>
A:Cross-references: DBJ:AP000060; NID:g5104188; PIDN:BAA79580.1; PID:d1043366; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0610

Query Match 2.0%; Score 8; DB 2; Length 376;

Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 299 ATALLTIS 306
|||||||
Db 26 ATALLTIS 33

RESULT 3
T08424
tweety protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08424
R:Malaszka, R.; de Couet, H.G.; Miklos, G.L.
Proc. Natl. Acad. Sci. U.S.A. 95, 3731-3736, 1998
A:Title: Data transferability from model organisms to human beings: insights from the fu
A:Reference number: Z16415; MUID:98188272
A:Accession: T08424
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-836 <MAL>
A:Cross-references: EMBL:AF017777; NID:93004652; PID:93004653
A:Experimental source: strain Cantons
C:Genetics:
A:Gene: tty
A:Introns: 46/3; 208/3; 463/2; 520/1; 524/1

Query Match 2.0%; Score 8; DB 2; Length 836;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 375 LGLLFYL 382
|||||||
Db 63 LGLLFYL 70

RESULT 4
C75397
hypothetical protein - *Deinococcus radiodurans* (strain R1)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: C75397
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250
A:Accession: C75397
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-67 <WHI>
A:Cross-references: GB:AE001987; GB:AE000513; NID:96459180; PID:AAF11003.1; PID:9645919
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1429
A:Map position: 1

Query Match 1.7%; Score 7; DB 2; Length 67;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 149 LMFVVI 155
|||||||
Db 12 LMFVVI 18

RESULT 5
S35174
cytochrome P450 (clone 9) - Madagascar periwinkle (fragment)

N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: *Catharantus roseus* (Madagascar periwinkle)
C:Date: 13-Jan-1995 #sequence_revision 09-May-1997 #text_change 11-Jun-1999
C:Accession: S35174
R:Meijer, A.H.; Souer, E.; Verpoorte, R.; Hoge, J.H.C.
Plant Mol. Biol. 22, 379-383, 1993
A:Title: Isolation of cytochrome P-450 cDNA clones from the higher plant *Catharantus*
A:Reference number: S35168; MUID:93283641
A:Accession: S35174
A:Molecule type: mRNA
A:Residues: 1-68 <MEI>
A:Cross-references: EMBL:X69781; NID:9395313; PID:CAA49436.1; PID:9395314
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreducta
E:2/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 1.7%; Score 7; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 SLSKQV 181
|||||||
Db 62 SLSKQV 68

RESULT 6
T15891
hypothetical protein D2096.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15891
R:Geisler, C.
Submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid D2096.
A:Reference number: Z18425
A:Accession: T15891
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-97 <GET>
A:Cross-references: EMBL:U40800; NID:91065946; PID:91065947; PID:AAA81489.1; CESP:ID2
C:Genetics:
A:Gene: CESP:D2096.5
A:Introns: 49/3; 84/2

Query Match 1.7%; Score 7; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 305 LSCGLST 311
|||||||
Db 34 LSCGLST 40

RESULT 7
W7MLHS
E7 protein - human papillomavirus type 16
C:Species: human papillomavirus type 16
C:Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 20-Aug-1999
C:Accession: A03688; S12367; T10428
R:Seedorf, K.; Krimer, G.; Durst, M.; Suhai, S.; Roweckamp, W.G.
Virology 145, 181-185, 1985
A:Title: Human papillomavirus type 16 DNA sequence.
A:Reference number: A22355; MUID:85246220
A:Accession: A03688
A:Molecule type: DNA
A:Residues: 1-98 <SEE>
A:Cross-references: GB:K02718; NID:9333031; PID:AAA46940.1; PID:9333033
R:Barbosa, M.S.; Edmonds, C.; Fisher, C.; Schiller, J.T.; Lowy, D.R.; Vousden, K.H.
EMBO J. 9, 153-160, 1990
A:Title: The region of the HPV E7 oncoprotein homologous to adenovirus E1a and SV40 1
A:Reference number: S12367; MUID:90107938

A:Accession: S12367
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-98 <BAR>
R:Kennedy, T.M.; Haddow, J.K.; Clements, J.B.
J. Virol. 65, 2093-2097, 1991
A:Title: A negative element in the human papillomavirus type 16 genome acts at the level
A:Reference number: Z17014; MUID:91162763
A:Accession: T10428
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-98 <REN>
A:Cross-references: EMBL:R02718; NID:g333031; PIDN:AAA46940.1; PID:g333033
C:Genetics:
A:Gene: E7
C:Superfamily: papillomavirus E7 protein
C:Keywords: DNA binding; early protein; transcription regulation; zinc finger
F:58-94/Region: zinc finger CCCC motif

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
|||||||
Db 27 QLNDSSE 33

RESULT 8
E71061
hypothetical protein PH187 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: E71061
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137
A:Accession: E71061
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-128 <KAW>
A:Cross-references: GB:AP000005; NID:g3326132; PIDN:AAA30287.1; PID:dl031230; PID:g32576
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by Genbank
C:Genetics:
A:Gene: PH187

Query Match 1.7%; Score 7; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 276 ILGSLPS 282
|||||||
Db 116 ILGSLPS 122

RESULT 9
T08514
outer membrane protein homolog - Enterobacter aerogenes plasmid R751
C:Species: Enterobacter aerogenes
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08514
R:Thorsted, P.B.; Shah, D.S.; Macartney, D.; Kostelidou, K.; Thomas, C.M.
Plasmid 36, 95-111, 1996
A:Title: Conservation of the genetic switch between replication and transfer genes of in
A:Reference number: Z16434; MUID:97118926
A:Accession: T08514
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-142 <THO>
A:Cross-references: EMBL:U67194; NID:g1572520; PID:g1572563
C:Genetics:
A:Gene: upf30.5
A:Genome: plasmid R751

Query Match 1.7%; Score 7; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 217 VVYIPRY 223
|||||||
Db 136 VVYIPRY 142

RESULT 10
H64325
hypothetical protein MJ0207 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Oct-1999
C:Accession: H64325
R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadov, P.W.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: H64325
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BUL>
A:Cross-references: GB:U67476; GB:L77117; NID:g1590942; PIDN:AMB98194.1; PID:g1498982
C:Genetics:
A:Map position: FOR198967-199422

Query Match 1.7%; Score 7; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GFISETL 131
|||||||
Db 122 GFISETL 128

RESULT 11
A45192
ribosomal protein S5 - Chlamydia trachomatis
C:Species: Chlamydia trachomatis
C:Date: 01-Jan-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A45192; F71505
R:Faul, R.; Gray, G.J.; Koehnke, N.R.; Gu, L.J.
J. Bacteriol. 174, 1205-1212, 1992
A:Title: Cloning and sequence analysis of the Chlamydia trachomatis spc ribosomal pro
A:Reference number: A42645; MUID:92138612
A:Accession: A45192
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-165 <KAU>
A:Cross-references: GB:M80325; NID:g144617; PIDN:AAA23178.1; PID:g144627
A:Note: sequence extracted from NCBI backbone (NCBIN:79464, NCBI:P:79474)
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: F71505
A:Molecule type: DNA
A:Residues: 1-165 <ARN>
A:Cross-references: GB:AE001323; GB:AE001273; NID:g3328931; PIDN:AAC68113.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:

A:Gene: r55
C:Superfamily: Escherichia coli ribosomal protein S5
C:Keywords: protein biosynthesis; ribosome

Query Match 1.7%: Score 7; DB 2; Length 165;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 301 ALTLTSC 307
|||||
Db 146 ALTLTSC 152

RESULT 12

S75094
hypothetical protein slr0271 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S75094
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.

A:Reference number: S74322; MUID:97061201

A:Accession: S75094

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-172 <KAN>

A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAAL7956.1; PID:d101868
C:Superfamily: Synechocystis hypothetical protein slr0271

Query Match 1.7%: Score 7; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 299 ATALLTL 305
|||||
Db 11 ATALLTL 17

RESULT 13

S56410
hypothetical 23.4K protein (vacb-aidB intergenic region) - Escherichia coli
N:Alternate names: hypothetical protein o212
C:Species: Escherichia coli

C>Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S56410; D65229
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.

Nucleic Acids Res. 23, 2105-2119, 1995

A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A:Reference number: S56314; MUID:95334362

A:Accession: S56410

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-212 <BUR>

A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97081.1; PID:g537026

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: D65229

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-212 <BLAT>

A:Cross-references: GB:AE000490; GB:U00096; NID:g2367356; PIDN:AACT7142.1; PID:g1790628;
A:Experimental source: strain K-12, substrain MG1655

C:Genetics:
A:Gene: yjfm

Query Match 1.7%: Score 7; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 350 VSGFLLS 356
|||||
Db 131 VSGFLLS 137

RESULT 14

S43291
FLT3/FLK2 ligand (clone r118) - mouse
C:Species: Mus musculus (house mouse)
C>Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999

C:Accession: S43291

R:Hannum, C.; Culpepper, J.; Campbell, D.; McClanahan, T.; Zurawski, S.; Bazan, J.F.;

felt, A.; Menich, M.; Kellner, G.; Nankikawa, R.; Rennick, D.; Roncarolo, M.G.; Zlotnik
Nature 368, 643-648, 1994

A:Title: Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of haematopo

A:Reference number: S43290; MUID:94195428

A:Accession: S43291

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-220 <HAN>

Query Match 1.7%: Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 TATALLT 304
|||||
Db 179 TATALLT 185

RESULT 15

I58343
flt3 ligand isoform 5H - mouse
C:Species: Mus sp. (mouse)
C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 05-Nov-1999

C:Accession: I58343

R:Lyman, S.D.; James, L.; Escobar, S.; Downey, H.; de Vries, P.; Brasel, K.; Stocking

Oncogene 10, 149-157, 1995

A:Title: Identification of soluble and membrane-bound isoforms of the murine flt3 lig

A:Reference number: I58343; MUID:95124710

A:Accession: I58343

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-220 <RES>

A:Cross-references: GB:S76459; NID:g913479; PIDN:AA33069.1; PID:g913480

Query Match 1.7%: Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 TATALLT 304
|||||
Db 179 TATALLT 185

Search completed: May 25, 2000, 14:57:56
Job time: 430 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:54:03 ; Search time 10.58 Seconds

(without alignments)
1154.295 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 401
Sequence: 1 MGVDFRLIPKVPKSLCSARY.....LIRGEADVQIMAKERKILTRL 401

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 83857 seqs, 30454973 residues

Word size : 0

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241	60.1	401	NP74_HUMAN	O00476 homo sapien
2	9	2.2	436	NP73_HUMAN	O00624 homo sapien
3	7	1.7	98	VE7_HPV16	P03129 human papill
4	7	1.7	151	Y207_METTL	O57660 methanococc
5	7	1.7	165	R55_CHL1R	P28543 chlamydia t
6	7	1.7	172	Y271_SYNY3	P73892 synechocyst
7	7	1.7	187	BACA_AZOB3	P39438 azospirillum
8	7	1.7	209	ATPO_DROME	O24439 drosophila
9	7	1.7	212	YJFM_ECOLI	P39295 escherichia
10	7	1.7	266	CB21_WHEAT	P04784 triticum ae
11	7	1.7	273	UCRI_MAIZE	P43727 zea mays (m
12	7	1.7	280	GCVK_HSV62	P52446 herpes slimp
13	7	1.7	305	SCRK_ECOLI	P40713 escherichia
14	7	1.7	324	LXB2_PHOUJ	P23147 photorhabdu
15	7	1.7	327	LXBI_PHOUJ	P19840 photorhabdu
16	7	1.7	379	ADH3_HORVU	P10848 hordeum vul
17	7	1.7	399	INBP_HUMAN	P49441 homo sapien
18	7	1.7	400	INBP_BOVIN	P21327 bos taurus
19	7	1.7	435	YN66_MYCTU	O05832 mycobacteri
20	7	1.7	465	NP71_HUMAN	O14916 homo sapien
21	7	1.7	465	NP71_MOUSE	O61983 mus musculu
22	7	1.7	465	NP71_RABIT	O28722 oryctolagus
23	7	1.7	465	NP71_RAT	O62795 rattus norv
24	7	1.7	562	GCVK_HSV60	P24446 herpes slimp
25	7	1.7	576	Y006_CAEEL	P34644 caenorhabdi
26	7	1.7	671	ALYS_ENTFA	P37710 enterococcu
27	7	1.7	835	Y422_MYCCE	P47661 mycoplasma
28	7	1.7	1597	RIR1_YEAST	P55552 saccharomyc
29	7	1.7	1804	YFAY_YEAST	P45583 saccharomyc
30	6	1.5	48	ATP8_SCHPO	P21536 schizosacch
31	6	1.5	49	LYS0_ECOLI	P02987 escherichia
32	6	1.5	60	MERC_PSEAE	P04139 pseudomonas
33	6	1.5	60	MERC_SHITL	P04337 shigella fl

34	6	1.5	64	1	RELX_RABER	P11952 raja erinac
35	6	1.5	101	1	CY02_CATRO	P25923 catharantnu
36	6	1.5	102	1	VE7_HPV36	P50811 human papill
37	6	1.5	102	1	YNI2_METTL	P05410 methanococc
38	6	1.5	103	1	VE7_HPV05	P06932 human papill
39	6	1.5	103	1	VE7_HPV08	P06430 human papill
40	6	1.5	103	1	VE7_HPV12	P36819 human papill
41	6	1.5	103	1	VE7_HPV47	P22423 human papill
42	6	1.5	103	1	VE7_HPV5B	P26559 human papill
43	6	1.5	108	1	IATP_PIG	O29307 sus scrofa
44	6	1.5	108	1	KVIV_HUMAN	P04430 homo sapien
45	6	1.5	115	1	TVC2_MOUSE	P03978 mus musculu

ALIGNMENTS

RESULT 1
ID NP74_HUMAN STANDARD; PRT: 401 AA.
AC 000476:
DT 15-JUL-1998 (Rel. 36, Created)
DR 15-JUL-1998 (Rel. 36, Last sequence update)
DI 15-JUL-1998 (Rel. 36, Last annotation update)
DE SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 4 (SODIUM/PHOSPHATE
DE COTRANSPORTER 4) (NA(+)/PI COTRANSPORTER 4).
GN SLC17A4 OR NP74.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP Ruddy D.A., Kronmal G.S., Lee V.K., Muntler G.A., Quintana L.,
RA Domingo R., Meyer N.C., Basava A., McJelland E., Fullan A.,
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C.,
RA Tsuchinashi Z., Wolff R.K., Schatzman R.C., Feder J.N.;
RL Submitted (JUN-1997) to the EMBL/Genbank/DDJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
CC INTO CELLS VIA NA+ COTRANSPORT (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
or send an email to license@isb-sb.ch).

CC EMBL: U90545; AAB53423.1;
DR Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
KW Sodium transport.
FT TRANSMEM 20 POTENTIAL.
FT TRANSMEM 108 128 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 365 386 POTENTIAL.
FT CARBOHYD 41 41 POTENTIAL.
FT CARBOHYD 49 49 POTENTIAL.
FT CARBOHYD 58 58 POTENTIAL.
FT CARBOHYD 58 58 POTENTIAL.
SQ SEQUENCE 401 AA: 44051 MW: 5235982091948285 CRC64;

Query Match 60.1%; Score 241; DB 1; Length 401;

Best Local Similarity 100.0%; Pred. No. 4.7e-235;

Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 SPWMTSKETKIYISSLKQVSSKOPLPKAMLRSLPTWISICIGCFSHQWVSTWVYI 220
DB 161 STFWMTSKETKIYISSLKQVSSKOPLPKAMLRSLPTWISICIGCFSHQWVSTWVYI 220

QY 221 PTYISSVYHVNIRDNGLSALPEIVAMVIGMYLADFLTKRRLTVRKIATITLGS 280
 DB 221 PTYISSVYHVNIRDNGLSALPEIVAMVIGMYLADFLTKRRLTVRKIATITLGS 280
 QY 281 PSSALIVSLPYINSGYITATALLTSCGSLTCSGSIYINVDIAPRYSFFLMGASRGFS 340
 DB 281 PSSALIVSLPYINSGYITATALLTSCGSLTCSGSIYINVDIAPRYSFFLMGASRGFS 340
 QY 341 STAPVIVPVSGFLSQDEFGFGRNVFLFFAVNLLGLFYILFEGADVOEWAKERKLT 400
 DB 341 STAPVIVPVSGFLSQDEFGFGRNVFLFFAVNLLGLFYILFEGADVOEWAKERKLT 400
 QY 401 L 401
 DB 401 L 401

RESULT 2
 NPT3_HUMAN STANDARD: PRT: 436 AA.
 ID NPT3_HUMAN 000624;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 3 (SODIUM/PHOSPHATE
 CO-TRANSPORTER 3) (NA(+)/PI CO-TRANSPORTER 3).
 GN SLC17A3 OR NPT3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RA Ruddy D.A., Kromal G.S., Lee V.K., Mintier G.A., Quintana L.,
 RA Domingo R., Meyer N.C., Basava A., McClelland E., Fulian A.,
 RA Mapa F.A., Moore T., Thomas W., Joeb D.B., Harmon C., Tsuchihashi Z.,
 RA Wolff R.K., Schatzman R.C., Feder J.N.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
 INTO CELLS VIA NA+ CO-TRANSPORT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U91328; AAB82085.1; -;
 DR EMBL; U90544; AAB53422.1; -;
 KW Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
 FT TRANSMEM 26 46
 FT TRANSMEM 98 118 POTENTIAL.
 FT TRANSMEM 130 150 POTENTIAL.
 FT TRANSMEM 183 203 POTENTIAL.
 FT TRANSMEM 211 231 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 317 337 POTENTIAL.
 FT TRANSMEM 386 406 POTENTIAL.
 FT CARBOHYD 47 47 POTENTIAL.
 FT CARBOHYD 56 56 POTENTIAL.
 FT CARBOHYD 68 68 POTENTIAL.
 FT CARBOHYD 69 69 POTENTIAL.
 SQ SEQUENCE 436 AA; 47350 MW; DF02F618E83A572F CRC64;

Query Match 2.2%; Score 9; DB 1; Length 436;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 GCVCCLMP 152
 DB 220 GCVCCLMP 228
 RESULT 3
 VE7_HPV16 STANDARD: PRT: 98 AA.
 ID VE7_HPV16
 AC P03129;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE E7 PROTEIN.
 GN E7.
 OS Human papillomavirus type 16.
 OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
 RN [1]
 RA SEQUENCE FROM N.A.
 RX MEDLINE: 85246220.
 RA Seedorf K., Kramer G., Durst M., Suhai S., Rowe& W.G.;
 RL "Human papillomavirus type 16 DNA sequence."; Virology 145:181-185(1985).
 RN [2]
 RA SEQUENCE FROM N.A.
 RX MEDLINE: 90218027.
 RA Schneider-Maunoury S., Pehau-Arnaudet G., Breitburd F., Orth G.;
 RL "Expression of the human papillomavirus type 16 genome in SK-V cells,
 a line derived from a vulvar intraepithelial neoplasia."; J. Gen. Virol. 71:809-817(1990).
 RN [3]
 RA SEQUENCE FROM N.A.
 RX Song Y.S., Kee S.H., Kim J.W., Park N.H., Kang S.B., Lee H.P.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RA SEQUENCE FROM N.A.
 RA Tornesello M.L., Buonaguro F.M., Meglio A., Buonaguro L.,
 RA Beth-Giraldo E., Giraldo G.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RA FUNCTION.
 RX MEDLINE: 88223347.
 RA Phelps W.C., Yee C.L., Munger K., Howley P.M.;
 RL "The human papillomavirus type 16 E7 gene encodes transactivation and
 transformation functions similar to those of adenovirus E1A."; Cell 53:539-547(1988).
 CC -1- FUNCTION: E7 PROTEIN HAS BOTH TRANSFORMING AND TRANS-ACTIVATING
 CC ACTIVITIES.
 CC -1- DISEASE: HPV16, IN COMPARISON TO HPV TYPES 6 AND 11, IS MORE
 CC OFTEN ASSOCIATED WITH MALIGNANT GENITAL CANCERS IN HUMANS.
 CC -1- SIMILARITY: LOCAL WITH ADENOVIRUS E1A AND SV40 LT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; K02718; AAA46940.1; -;
 DR EMBL; D00735; BAA00633.1; -;
 DR EMBL; U76411; AAB18962.1; -;
 DR EMBL; U76412; AAB18963.1; -;
 DR EMBL; U76413; AAB18964.1; -;
 DR EMBL; AF003020; AAB70737.1; -;
 DR EMBL; AF003023; AAB70740.1; -;
 DR EMBL; AF003024; AAB70741.1; -;
 DR EMBL; AF003025; AAB70742.1; -;
 DR EMBL; AF003026; AAB70743.1; -;
 DR PTR; A03688; W7WLS.
 DR PRAM; PF00527; E7; 1.
 KW Early protein; Transcription regulation; Transforming protein;
 DNA-binding; Trans-acting factor.

FT DOMAIN 58 61 C-XX-C MOTIF.
SO SEQUENCE 98 AA; 11022 MW; 9BD612534CD2C9EB CRC64;
Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 QLNDSSE 62
DB 27 QLNDSSE 33
RESULT 4
Y207_METJA
ID Y207_METJA STANDARD; PRT; 151 AA.
AC 057660;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0207.
GN MJ0207.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
RA Klenk H.-P., Frazer C.M., Smith H.O., Moese C.R., Venter J.C.;
RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.*;
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: WEAK, TO M.JANNASCHII MJ1342 AND TO THE N-TERMINAL OF
M.JANNASCHII MJ1450.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D67476; AAB98194.1; -.
DR TIGR; MJ0207; -.
KW Hypothetical protein.
SO SEQUENCE 151 AA; 16912 MW; 03F78A3F2EBCA189 CRC64;
Query Match 1.7%; Score 7; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 125 GFISETL 131
DB 122 GFISETL 128
RESULT 5
RS5_CHLTR
ID RS5_CHLTR STANDARD; PRT; 165 AA.
AC P28543;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S5.
GN RPS5 OR RS5 OR CTS12.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-L2/434/BU;
RX MEDLINE; 92138612.
RA Kaul R., Gray G.J., Koehncke N.R., Gu L.J.;
RT "Cloning and sequence analysis of the Chlamydia trachomatis spc
RT ribosomal protein gene cluster.";
RL J. Bacteriol. 174:1205-1212(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE; 99000809.
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -1- FUNCTION: PROTEIN S5 IS IMPORTANT IN THE ASSEMBLY AND FUNCTION OF
CC THE 30S RIBOSOMAL SUBUNIT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M80325; AAA23178.1; -.
DR EMBL; AE001323; AAC68113.1; -.
DR PIR; A45192; A45192.
DR HSSP; P02357; 1PKP.
DR PROSITE; PS00585; RIBOSOMAL_S5; 1.
DR PFAM; PF00333; Ribosomal_S5; 1.
KW Ribosomal protein.
SO SEQUENCE 165 AA; 17762 MW; 66F3E0AA8409481A CRC64;
Query Match 1.7%; Score 7; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 301 ALUTLSC 307
DB 146 ALUTLSC 152
RESULT 6
Y271_SYNY3
ID Y271_SYNY3 STANDARD; PRT; 172 AA.
AC P73892;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 19.4 KD PROTEIN SLR0271 PRECURSOR.
GN SLR0271.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Cyrococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyaajima N., Hirasawa M., Sugiyura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuo A., Mureki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90910; BAA17956.1; -
KW Hypothetical protein; signal.
FT SIGNAL 1 24
FT CHAIN 1 24
SQ SEQUENCE 172 AA: 19439 MW: 2606EB831BD440CA CRC64;

Query Match 1.7%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 299 ATALLTL 305
Db 11 ATALLTL 17

RESULT 7
BACA_AZOBR STANDARD; PRT; 187 AA.
AC P39438;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE BACTIRACIN RESISTANCE PROTEIN HOMOLOG (PUTATIVE UNDECAPRENOL KINASE)
DE (EC 2.7.1.66) (FRAGMENT).
OS Azospirillum brasilense.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC Azospirillum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SP;
RX MEDLINE; 93155143.
RA Pelanda R., Vanoni M.A., Perego M., Pinbelli L., Galizzi A.,
RA Curti B., Zanetti G.;
RT "Glutamate synthase genes of the diazotroph *Azospirillum brasilense*.
RT Cloning, sequencing, and analysis of functional domains.";
RL J. Biol. Chem. 268:3099-3106(1993).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (NOV-1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: TO E.COLI BACA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L04300; -; NOT_ANNOTATED_CDS.
KW Antibiotic resistance; Transmembrane; Transferase; Kinase.
FT TRANSMEM 13 33
FT TRANSMEM 45 65
FT TRANSMEM 85 105
FT TRANSMEM 108 128
FT TRANSMEM 144 164
FT NON_TER 187 187

SQ SEQUENCE 187 AA: 19842 MW: E04C8995BCFDB63A CRC64;

Query Match 1.7%; Score 7; DB 1; Length 187;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 274 ATILGSL 280
Db 169 ATILGSL 175

RESULT 8
ATPO_DROME STANDARD; PRT; 209 AA.
AC Q24439;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE ATP SYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEIN PRECURSOR,
DE MITOCHONDRIAL (EC 3.6.1.34) (OSCP).
OS OSCP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE; 99168769.
RA Caggese C., Ragone G., Perrini B., Moschetti R., De Pinto V.,
RA Calzati R., Barsanti P.;
RT "Identification of nuclear genes encoding mitochondrial proteins:
RT isolation of a collection of D. melanogaster cDNAs homologous to
RT sequences in the Human Gene Index database.";
RL Mol. Gen. Genet. 261:64-70(1999).
CC -1- FUNCTION: THIS PROTEIN SEEMS TO BE PART OF THE STALK THAT LINKS
CC CF(0) TO CF(1). IT EITHER TRANSMITS CONFORMATIONAL CHANGES FROM
CC CF(0) INTO CF(1) OR IS IMPLICATED IN PROTON CONDUCTION.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SIMILARITY: BELONGS TO THE ATPASE DELTA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X99666; CAA67980.1; -
DR FLVBASE; FB94001691; OSCP.
DR PROSITE; PS00389; ATPASE_DELTA; 1.
DR PFAM; PF00213; OSCP; 1.
KW Hydroxylase; ATP synthetase; CF(1); Hydrogen ion transport;
KW Mitochondrion; Transil peptide.
FT TRANSIT 1 209
FT CHAIN 1 209
SQ SEQUENCE 209 AA: 22421 MW: 5004018524DFCAB2 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 VNILGL 379
Db 102 VNILGL 108

```

RESULT 9
VFPM_ECOLI
ID VFPM_ECOLI STANDARD: PRT: 212 AA.
AC P39295;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 23.4 KD PROTEIN IN RNR-AIDB INTERGENIC REGION.
GN VFPM.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 95334362.
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14003; AAA97081.1; -.
DR EMBL: AE000490; AAC77142.1; -.
DR ECGENE: EG12488; yjfm.
KM Hypothetical protein.
SQ SEQUENCE 212 AA; 23352 MW; 0C69332DA15AEAE4 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 350 VSGFLLS 356
DB 131 VSGFLLS 137

RESULT 10
CB21_WHEAT
ID CB21_WHEAT STANDARD: PRT: 266 AA.
AC P04784;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CHLOROCHYL A-B BINDING PROTEIN PRECURSOR (LHCII TYPE I CAB) (LHCP).
GN WHA1.6.
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Triticum.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85295969.
RA Lampka G.K., Morelli G., Chua N.-H.;
RT "Structure and developmental regulation of a wheat gene encoding the
RT major chlorophyll a/b-binding polypeptide.";
RL Mol. Cell. Biol. 5:1370-1378(1985).
CC -----
CC -1- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT
CC RECEPTOR, IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS
CC WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN
CC EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF
CC GRANAL MEMBRANES AND PHOREGULATED BY REVERSIBLE PHOSPHORYLATION
CC OF ITS THREONINE RESIDUES; BOTH ARE BELIEVED TO MEDATE THE

```

```

CC -----
CC DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.
CC -1- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS (A & B) AND
CC CHLOROCHYL A-B BINDING PROTEINS.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M10144; AAA34260.1; -.
DR PIR: A23755; CDWT.
DR PFAM: PF00504; chloroa_b-bind; 1.
KM Chlorophyll; Photosynthesis; Photosystem I; Photosystem II;
KM Thylakoid membrane; Chloroplast; Transit peptide; Multigene family;
KM Transmembrane; Phosphorylation.
FT TRANSIT 1 33 CHLOROPLAST (PROBABLE).
FT CHAIN 34 266 CHLOROCHYL A-B BINDING PROTEIN.
FT TRANSMEM 100 119 POTENTIAL.
FT TRANSMEM 152 172 POTENTIAL.
FT TRANSMEM 220 236 POTENTIAL.
SQ SEQUENCE 266 AA; 28264 MW; 9D5719759DF3C4E8 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 266;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 LPSSALI 286
DB 21 LPSSALI 27

RESULT 11
UCRL_MAIZE
ID UCRL_MAIZE STANDARD: PRT: 273 AA.
AC P49727;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE UBIOUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT, MITOCHONDRIAL
DE PRECURSOR (EC 1.10.2.2) (RIESKE IRON-SULFUR PROTEIN) (RISP).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92073358.
RA Huang J.T., Struck F., Matzinger D.F., Levings C.S. III;
RT "Functional analysis in yeast of cDNA coding for the mitochondrial
RT Rieske iron-sulfur protein of higher plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10716-10720(1991).
CC -----
CC -1- FUNCTION: COMPONENT OF THE UBIOUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C = O +
CC 2 FERROCYTOCHROME C.
CC -1- SUBUNIT: BCI COMPLEX CONTAINS 10 SUBUNITS; 3 RESPIRATORY
CC SUBUNITS, 2 CORE PROTEINS AND 5 LOW-MOLECULAR WEIGHT PROTEINS.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC -1- MISCELLANEOUS: THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S
CC PROTEIN.
CC -----
CC -1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
CC BACTERIAL, CHLOROPLAST).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M77224; AAA33507.1; -
DR HSSP: P13272; 1R1E.
DR MAI2EDB: 30151; -
DR PRINTS: PR00162; RIESKE.
DR PROSITE: PS00199; RIESKE_1; 1.
DR PROSITE: PS00200; RIESKE_2; 1.
DR PFMW: PF00355; Rieske; 1.
KW Mitochondrion; Electron transport; Respiratory chain; Iron-sulfur;
KW Oxidoreductase; Inner membrane; Transmembrane; Transit peptide.
FT TRANSIT 1 61 MITOCHONDRION (POTENTIAL).
FT CHAIN 62 273 UBIQUINOL-CYTOCHROME C REDUCTASE IRON-
FT METAL 216 216 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 218 218 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 235 235 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 238 238 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT DISULFD 221 237 BY SIMILARITY.
SQ SEQUENCE 273 AA; 29835 MW; 43B08C1169B1A2C CRC64;

Query Match 1.7%; Score 7; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 335 ASRGFS 341
|||||||
DB 56 ASRGFS 62

RESULT 12
GCVK_HSV62 STANDARD; PRT; 280 AA.
AC P52446;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE POSSIBLE GANCICLOVIR KINASE (EC 2.7.1.1) (FRAGMENT).
GN U69 OR CH2R.
OS Herpes simplex virus (type 6 / strain 229).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96195263.
RA Lindquister G.J., Inoue N., Allen R.D., Castelli J.W.,
RA Stamey F.R., Dambaugh T.R., O'Brian J.J., Danovich R.M.,
RA Frankel N., Pellett P.E.;
RT "Restriction endonuclease mapping and molecular cloning of the human
RT herpesvirus 6 variant B strain 229 genome.";
RL Arch. Virol. 141:367-379(1996).
CC -1- FUNCTION: PHOSPHORYLATES THE ANTIVIRAL NUCLEOSIDE ANALOG
CC GANCICLOVIR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HCMV UL97 / HVS 36 FAMILY.
CC -1- SIMILARITY: TO TYROSINE-PROTEIN KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L14772; AAB06352.1; -
DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN KINASE_DOM; FALSE_NEG.
KW Transferase; Kinase; ATP-binding.

FT NON_TER 1 1 BY SIMILARITY.
FT ACT_SITE 31 31
SQ SEQUENCE 280 AA; 32348 MW; 930E2ED2A50A9107 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 375 LIGLIFY 381
|||||||
DB 190 LIGLIFY 196

RESULT 13
SCRK_ECOLI STANDARD; PRT; 305 AA.
AC P40713;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE FRUCTOKINASE (EC 2.7.1.4).
GN CSCK.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -1- SIMILARITY: BELONGS TO THE PKB FAMILY OF CARBOHYDRATE KINASES.
CC -1- SIMILARITY: BELONGS TO THE PKB FAMILY OF CARBOHYDRATE KINASES.
CC -1- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE = ADP + D-FRUCTOSE
CC 6-PHOSPHATE.
CC
CC EMBL: X81461; CA57218.1; -
DR PROSITE: PS00583; PKB_KINASES_1; 1.
DR PROSITE: PS00584; PKB_KINASES_2; 1.
DR PFMW: PF00294; pfb; 1.
KW Transferase; Kinase.
SQ SEQUENCE 305 AA; 33081 MW; 5F9C8D59B66C3E55 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 305;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 LOSTALS 111
|||||||
DB 125 LOSTALS 131

RESULT 14
LXB2_PHOLU STANDARD; PRT; 324 AA.
AC P23147;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ALKANAL MONOOXYGENASE BETA CHAIN (EC 1.14.14.3) (BACTERIAL LUCIFERASE
DE BETA CHAIN).
GN LUXB.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterius.

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HW:
RX MEDLINE: 91139581.
RA "Cloning and nucleotide sequences of lux genes and characterization
RT of luciferase of Xenorhabdus luminescens from a human wound.";
RL J. Bacteriol. 173:1399-1405(1991).
CC -1- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA. THE
CC SPECIFIC ROLE OF THE BETA SUBUNIT IS UNKNOWN, BUT IT IS ABSOLUTELY
CC REQUIRED FOR BIOLUMINESCENCE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: FMNH(2) + RCHO + O(2) -> FMN + RCOOH + H(2)O
CC + LIGHT.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M62917; AAA63566.1; -.
DR PIR: D38448; D38448.
DR HSSP: P07739; 1XKJ.
DR PRINTS: PR00089; LUCIFERASE.
DR PROSITE: PS00494; BACTERIAL_LUCIFERASE; 1.
DR PFAM: PF00296; bac_luciferase; 1.
DR Photoprotein; Luminescence; Oxidoreductase; Monoxygenase;
KW Flavoprotein; FMN.
SQ SEQUENCE 324 AA; 37115 MW; E7D81324D344A975 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 324;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 349 TVSGFLL 355
DB 57 TVSGFLL 63

RESULT 15
LXBL_PHOTOU STANDARD; PRT; 327 AA.
AC P19840;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE ALKALAL MONOOXYGENASE BETA CHAIN (EC 1.14.14.3) (BACTERIAL LUCIFERASE
DE BETA CHAIN).
GN LUXB.
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photorhabdus.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-ATCC 29999;
RX MEDLINE: 90375532.
RA Sziltnier R., Meighen E.;
RT "Nucleotide sequence, expression, and properties of luciferase coded
RT by lux genes from a terrestrial bacterium.";
RL J. Biol. Chem. 265:16581-16587(1990).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN-ATCC 29999;
RX MEDLINE: 92355513.
RA Meighen E.A., Sziltnier R.B.;
RT "Multiple repetitive elements and organization of the lux operons of
RT luminescent terrestrial bacteria.";
RL J. Bacteriol. 174:5371-5381(1992).
RN [3]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN-HW:
RX MEDLINE: 90343746.
RA Johnston T.C., Rucker E.B., Cochrum L., Hruska K.S., Vandegrift V.;
RT "The nucleotide sequence of the luxA and luxB genes of Xenorhabdus
RT luminescens HM and a comparison of the amino acid sequences of
RT luciferases from four species of bioluminescent bacteria.";
RL Biochem. Biophys. Res. Commun. 170:407-415(1990).
CC -1- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA. THE
CC SPECIFIC ROLE OF THE BETA SUBUNIT IS UNKNOWN, BUT IT IS ABSOLUTELY
CC REQUIRED FOR BIOLUMINESCENCE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: FMNH(2) + RCHO + O(2) -> FMN + RCOOH + H(2)O
CC + LIGHT.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M57416; AAA27624.1; -.
DR DR EMBL: M90093; AAA27620.1; -.
DR DR EMBL: M55977; AAA27627.1; -.
DR PIR: C37898; C37898.
DR HSSP: P07739; 1XKJ.
DR PRINTS: PR00089; LUCIFERASE.
DR PROSITE: PS00494; BACTERIAL_LUCIFERASE; 1.
DR PFAM: PF00296; bac_luciferase; 1.
DR Photoprotein; Luminescence; Oxidoreductase; Monoxygenase;
KW Flavoprotein; FMN.
FT VARIANT 85 A -> R (IN STRAIN HM).
SQ SEQUENCE 327 AA; 37595 MW; 1905AF01658BE56E CRC64;

Query Match 1.7%; Score 7; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 349 TVSGFLL 355
DB 57 TVSGFLL 63
```

Search completed: May 25, 2000, 14:59:49
Job time: 346 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:52:26 ; Search time 17.37 Seconds
(without alignments)
1600.632 Million cell updates/sec

Title: US-09-391-958-1
Perfect score: 401
Sequence: 1 MOVDELIPKRVPSICSAKY.....LIFGEADVQEWAKRKILRL 401

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 225878 seqs, 69334122 residues

Word size : 0

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_12:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.0	333	2 09Z6L1	09Z6L1 Chlamydia p
2	8	2.0	336	12 09WR6	09WR6 macaca mula
3	8	2.0	376	1 09TEG6	09TEG6 aetopyrum p
4	8	2.0	497	4 09Y2C5	09Y2C5 homo sapien
5	8	2.0	836	5 061343	061343 drosophila
6	7	1.7	49	9 038059	038059 bacterioph
7	7	1.7	63	4 015140	015140 homo sapien
8	7	1.7	68	10 008570	008570 catharant
9	7	1.7	97	5 019002	019002 caenorhabd
10	7	1.7	98	12 012338	012338 human papil
11	7	1.7	128	1 058910	058910 pyrococcus
12	7	1.7	141	2 066368	066368 unidentified
13	7	1.7	142	2 P71192	P71192 escherichia
14	7	1.7	209	5 024439	024439 drosophila
15	7	1.7	235	2 09W253	09W253 thermocoga
16	7	1.7	248	8 09XLY0	09XLY0 hepiatius ka
17	7	1.7	249	7 008607	008607 cyphotliapl
18	7	1.7	257	2 052995	052995 rhizobium m
19	7	1.7	264	5 097455	097455 caenorhabd

20	7	1.7	283	2 006523	006523 haemophilus
21	7	1.7	283	2 087121	087121 actinobacil
22	7	1.7	283	2 09ZAB7	09ZAB7 actinobacil
23	7	1.7	296	5 021199	021199 caenorhabd
24	7	1.7	299	2 088084	088084 enterococcu
25	7	1.7	304	13 091580	091580 xenopus lae
26	7	1.7	310	2 067396	067396 aquifex aeo
27	7	1.7	320	2 052533	052533 pseudomonas
28	7	1.7	324	2 056822	056822 xenorhabdus
29	7	1.7	344	4 000421	000421 homo sapien
30	7	1.7	344	6 09XSD7	09XSD7 macaca mula
31	7	1.7	356	4 075307	075307 homo sapien
32	7	1.7	480	2 054442	054442 serralia ma
33	7	1.7	480	2 09WMD3	09WMD3 serralia ma
34	7	1.7	484	4 013793	013793 homo sapien
35	7	1.7	492	4 P96820	P96820 mycobacteri
36	7	1.7	512	3 051750	051750 borrelia bu
37	7	1.7	512	2 013985	013985 schizosacch
38	7	1.7	512	10 082390	082390 arabidopsis
39	7	1.7	527	11 P70486	P70486 rattus norv
40	7	1.7	534	4 014611	014611 homo sapien
41	7	1.7	534	11 09WVK4	09WVK4 mus musculi
42	7	1.7	537	5 018274	018274 caenorhabd
43	7	1.7	554	10 024298	024298 plasmu sativ
44	7	1.7	554	10 065197	065197 plasmu sativ
45	7	1.7	560	11 062634	062634 rattus norv

ALIGNMENTS

RESULT 1
ID 09Z6L1 PRELIMINARY; PRT; 333 AA.
AC 09Z6L1;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE ASPARTATE DEHYDROGENASE.
GN ASD.
OS Chlamydia pneumoniae.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWLD29;
RA KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., OLINGER L.,
RA GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;
RT "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO01685; AAD19185.1; -.
SQ SEQUENCE 333 AA; 37300 MW; 0C9213A4 CRC32;

Query Match 2.0%; Score 8; DB 2; Length 333;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 182 GSSKOPLP 189
Db 206 GSSKOPLP 213
RESULT 2
ID 09WR9 PRELIMINARY; PRT; 336 AA.
AC 09WR9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HYPOTHETICAL 36.9 KD PROTEIN.
OS Macaca mulatta rhadinovirus 17577.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 99174001.
 RA SEALES R.P., BERGOUAM E.P., AXTHELM M.K., WONG S.W.;
 RT "Sequence and genomic analysis of a Rhesus macaque rhadinovirus with
 RT similarity to Kaposi's sarcoma-associated herpesvirus/human
 RT herpesvirus 8";
 RL J. Virol. 73:3040-3053(1999).
 DR EMBL: AF083501; NAD21359.1; -.
 KM Hypothetical protein.
 SQ SEQUENCE 336 AA; 36913 MW; EC16FACE CRC32;

Query Match 2.0%; Score 8; DB 12; Length 336;
 Best Local Similarity 100.0%; Pred. No. 8.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 373 VNLGLLF 380
 Db 258 VNLGLLF 265

RESULT 3
 Q9YEG6 PRELIMINARY; PRT; 376 AA.
 AC Q9YEG6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE 376AA LONG HYPOTHEICAL PROTEIN.
 GN APE0610.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Aeropyrum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K1;
 RX MEDLINE: 99310339.
 RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
 RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
 RA HOSOGIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
 RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
 RA NOMURA N., SAKO Y., KIRUCHI H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000060; BAA79580.1; -.
 SQ SEQUENCE 376 AA; 39239 MW; 775B0176 CRC32;

Query Match 2.0%; Score 8; DB 1; Length 376;
 Best Local Similarity 100.0%; Pred. No. 9.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 299 ATALITLS 306
 Db 26 ATALITLS 33

RESULT 4
 Q9YZC5 PRELIMINARY; PRT; 497 AA.
 AC Q9YZC5;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE NA/PO4 COTRANSPORTER HOMOLOG.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 99253143.

RA SHIBUI A., TSUNODA T., SEKI N., SUZUKI Y., SUGANE K., SUGANO S.;
 RT "Isolation and chromosomal mapping of a novel human gene showing
 RT homology to Na+/PO4 cotransporter";
 RL J. Hum Genet. 44:190-192(1999).
 DR EMBL: AB020527; BAA76653.1; -.
 SQ SEQUENCE 497 AA; 54055 MW; 128F47B4 CRC32;

Query Match 2.0%; Score 8; DB 4; Length 497;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 EFGWRNVF 367
 Db 456 EFGWRNVF 463

RESULT 5
 O61343 PRELIMINARY; PRT; 836 AA.
 ID O61343;
 AC O61343;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE TWERTY.
 GN TTY.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTONS;
 RX MEDLINE: 96129280.
 RA DE COUET H.G., FONG K.S., WEEDS A.G., MCLAUGHLIN P.J., MIKIOS G.L.;
 RT "Molecular and mutational analysis of a gelsoilin-family member encoded
 RT by the flightless I gene of Drosophila melanogaster";
 RL Genetics 141:1049-1059(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTONS;
 RX MEDLINE: 97289742.
 RA MIKIOS G.L., YAMAMOTO M., BURNS R.G., MALESZKA R.;
 RT "An essential cell division gene of Drosophila, absent from
 RT Saccharomyces, encodes an unusual protein with tubulin-like and
 RT myosin-like peptide motifs";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:5189-5194(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTONS;
 RA MALESZKA R., DE COUET H.G., MIKIOS G.L.G.;
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(1998).
 DR EMBL: AF017777; AAC28400.1; -.
 DR EMBL: FB90015558; tly.
 SQ SEQUENCE 836 AA; 92838 MW; FE221871 CRC32;

Query Match 2.0%; Score 8; DB 5; Length 836;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 LGLGLFYL 382
 Db 63 LGLGLFYL 70

RESULT 6
 Q38059 PRELIMINARY; PRT; 49 AA.
 ID Q38059;
 AC Q38059;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE HYPOTHETICAL 5.7 KD PROTEIN.
OS Bacteriophage PZA.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86056991.
RA PACES V., VICEK C., URBANEK P., HOSTOMSKY Z.;
RT "Nucleotide sequence of the major early region of Bacillus subtilis
phage PZA, a close relative of phl 29."
RL Gene 38:45-56(1985).
DR EMBL: M1813; AAA8482.1; -
KM Hypothetical protein.
SQ SEQUENCE 49 AA; 5650 MW; FDI8C640 CRC32;

Query Match 1.7%; Score 7; DB 9; Length 49;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 SGITINV 321
Db 29 SGITINV 35

RESULT 7
ID 015140 PRELIMINARY; PRT: 63 AA.
AC 015140;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE NA.K-ATPASE BETA-3 SUBUNIT PSEUDOGENE, COMPLETE SEQUENCE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eukaryota; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA MALIK N., CAMPBELL V., SANCHEZ-WATTS G., WATTS A., SCHERR S.,
RA BEATTY B., GROS P., LEVENSON R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF005898; AAB61675.1; -
DR PFMW: PF00287; Na.K-ATPase; 1.
SQ SEQUENCE 63 AA; 7591 MW; 4F685A8E CRC32;

Query Match 1.7%; Score 7; DB 4; Length 63;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 378 LFFYLIF 384
Db 37 LFFYLIF 43

RESULT 8
ID 008570 PRELIMINARY; PRT: 68 AA.
AC 008570;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CYTOCHROME P-450 (EC 1.14.14.1) (CR9) (FRAGMENT).
GN P450CR9.
OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Gentianales; Apocynaceae;
OC Catharanthus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-(L.) G.DON;
RX MEDLINE: 93283641.
RA MEIJER A.H., SOUER E., VERPOORTE R., HOGE J.H.C.;

RT "Isolation of cytochrome P-450 cDNA clones from the higher plant
Catharanthus roseus by a PCR strategy."
RL Plant Mol. Biol. 22:379-383(1993).
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
MONOOXYGENASES. MAY BE A GERANIOL-10-HYDROXYLASE.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- PATHWAY: INDOLE ALKALOID SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE BOUND.
CC -1- SIMILARITY: MEMBER OF THE CYTOCHROME P-450 FAMILY.
DR EMBL: X69781; CAA49436.1; -
DR MENDEL: 11599; Catro:113;11599.
KM Oxidoreductase; Monooxygenase; Electron transport; Transmembrane;
KW Heme.
FT NON_TER 1 1
SQ SEQUENCE 68 AA; 7579 MW; 752F8C06 CRC32;

Query Match 1.7%; Score 7; DB 10; Length 68;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 175 SSLKQV 181
Db 62 SSLKQV 68

RESULT 9
ID 019002 PRELIMINARY; PRT: 97 AA.
AC 019002;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE COSMID D2096.
GN D2096.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
OC Rhabdita; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BRKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA GEISEL C.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U40800; AAA81489.1; -
SQ SEQUENCE 97 AA; 11281 MW; E5A676F5 CRC32;

Query Match 1.7%; Score 7; DB 5; Length 97;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 305 LSCGLST 311
Db 305 LSCGLST 311

DB 34 LSCGLST 40

```
RESULT 10
ID 012338 PRELIMINARY; PRT; 98 AA.
AC 012338;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE E7 PROTEIN.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PC-4;
RX MEDLINE; 97437474.
RA TORRESIELLO M.L., BUONAGURO F.M., MEGLIO A., BUONAGURO L.,
RA BERTH-GIRALDO E., GIRALDO G.;
RT "Sequence variations and viral genomic state of human papillomavirus
RT type 16 in penile carcinomas from Ugandan patients.";
RL J. Gen. Virol. 78:0-0(0).
DR EMBL; AF003022; AAB70739.1; -.
DR PFMW; PF00527; E7; 1.
SQ SEQUENCE 98 AA; 10969 MW; 2D8CB74C CRC32;
```

Query Match 1.7%; Score 7; DB 12; Length 98;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QUNDSE 62
DB 27 QUNDSE 33

```
RESULT 11
ID 058910 PRELIMINARY; PRT; 128 AA.
AC 058910;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, last annotation update)
DE 128AA LONG HYPOTHETICAL PROTEIN.
GN PH1187.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SETINE M., BABA S., KOSUGI H., HOSOMIYA A., NGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMURA M., OHFUKU Y.,
RA FUJAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000005; BAA30287.1; -.
SQ SEQUENCE 128 AA; 14403 MW; 0E4F59C0 CRC32;
```

Query Match 1.7%; Score 7; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 ILGSLPS 282
DB 116 ILGSLPS 122

RESULT 12

```
ID 066368 PRELIMINARY; PRT; 141 AA.
AC 066368;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN NIFH.
OS unidentified nitrogen-fixing bacteria.
OC Bacteria.
RN [1]
RP SEQUENCE FROM N.A.
RC OKUMA M., NODA S., KUDO T.;
RA Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB011958; BAA28491.1; -.
DR HSSP; P00456; ICR2.
DR PROSITE; PS00692; NIFH_FRXC_2; 1.
DR PROSITE; PS00746; NIFH_FRXC_1; 1.
DR PFMW; PF00142; fer4_NifH; 1.
FT NON_TER 141
FT NON_TER 141
SQ SEQUENCE 141 AA; 15075 MW; 100DF20C CRC32;
```

Query Match 1.7%; Score 7; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 FGVGCV 146
DB 64 FGVGCV 70

```
RESULT 13
ID P71192 PRELIMINARY; PRT; 142 AA.
AC P71192;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, last annotation update)
DE UPF30.5.
GN UPF30.5.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC SMITH C.A., THOMAS C.M.;
RX MEDLINE; 85215496.
RA "Comparison of the nucleotide sequences of the vegetative replication
RT origins of broad host range incp plasmids R751 and RK2 reveals
RT conserved features of probable functional importance.";
RL Nucleic Acids Res. 13:557-572(1985).
DR EMBL; U67194; AAC64458.1; -.
KW Plasmid.
SQ SEQUENCE 142 AA; 15017 MW; E95FFD8A CRC32;
```

Query Match 1.7%; Score 7; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 VVYIPY 223
DB 136 VVYIPY 142

```
RESULT 14
ID Q24439 PRELIMINARY; PRT; 209 AA.
AC Q24439;
```

01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
 DE ATP SYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEIN PRECURSOR,
 GN MITOCHONDRIAL (EC 3.6.1.34) (OSCP).
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyridiidae; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE: 99168769.
 RA CAGGSE C., RAGONE G., PERRINI B., MOSCHETTI R., DE PINTO V.,
 RA CAIZZI R., BASANTI P.;
 RT Identification of nuclear genes encoding mitochondrial proteins:
 RT Isolation of a collection of D. melanogaster cDNAs homologous to
 RT sequences in the Human gene index database.";
 RL Mol. Gen. Genet. 261:64-70(1999)
 CC -1- FUNCTION: THIS PROTEIN SEEMS TO BE PART OF THE STALK THAT LINKS
 CC CF(0) TO CF(1). IT EITHER TRANSMITS CONFORMATIONAL CHANGES FROM
 CC CF(0) INTO CF(1) OR IS IMPLICATED IN PROTON CONDUCTION.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE DELTA CHAIN FAMILY.
 DR EMBL: X99666; CAA67980.1; -;
 DR FLYBASE: FBgn0016691; OSCP.
 DR PROSITE: PS00389; ATPASE_DELTA. 1.
 DR PRAM: PF00213; OSCP. 1.
 DR PRINTS: PR00125; ATPASEDELTA.
 KW Hydrolyase; ATP synthesis; CF(1); Hydrogen ion transport;
 KW Mitochondrion; Transist peptide.
 FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
 FT CHAIN ? 209 ATP SYNTHASE OLIGOMYCIN SENSITIVITY
 FT CONFERRAL PROTEIN.
 SQ SEQUENCE 209 AA; 22421 MW; 444323B3 CRC32;

Query Match 1.7%; Score 7; DB 5; Length 209;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 VNLGLL 379
 DB 102 VNLGLL 108

RESULT 15
 O9W253
 ID O9W253 PRELIMINARY; PRT; 235 AA.
 AC O9W253;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN TM0584.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 99287316.
 RA NELSON K.E., CLAYTON R.A., GILL S.R., GRINN M.L., DODSON R.J.,
 RA HAFI D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
 RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
 RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
 RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
 RA SMITH H.O., VENTER J.C., FRASER C.M.;
 RT "Evidence for lateral gene transfer between Archaea and bacteria from
 RT genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).

[2]
 RN SEQUENCE FROM N.A.
 RP NELSON K.E., CLAYTON R.A., GILL S.R., GRINN M.L., DODSON R.J.,
 RA HAFI D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
 RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
 RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
 RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
 RA SMITH H.O., VENTER J.C., FRASER C.M.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE001733; AAD35669.1; -;
 SQ SEQUENCE 235 AA; 26394 MW; 50FA4BFA CRC32;

Query Match 1.7%; Score 7; DB 2; Length 235;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 110 LSGMLG 116
 DB 180 LSGMLG 186

Search completed: May 25, 2000, 14:59:05
 Job time: 399 sec

OM of: US-09-391-958-1 to: EST:* out-format: pfs
 Date: May 25, 2000 3:10 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frimex.p2n.model -DEF=x1p
 -Q/cgr2_1/USFTO.spool/US09391958/rmat_24052000_213605_816/app-query.fasta.1
 -DB=EST -QW=fastap -SUFFIX=oligo.rst -GAPOP=4.500
 -GAPOP=4.500 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -OGAPOP=0.050 -OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFM=pfs -NORM=ext -MTLEN=0
 -MAXLEN=1000000 -USER=US09391958 -NCPU=6 -ICPU=3 -NO_XLFPX -WAIT
 -THREADS=1

Search information block:
 Query: US-09-391-958-1
 Query length: 401
 Database: EST:*
 Database sequences: 4857316
 Database length: 202661650
 Search time (sec): 602.260000

WARN: XGAPOP and YGAPEXT must be equal. Assuming YGAPOP=XGAPOP=60.000
 WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Sequence	Strd	Orig	ZScore	EScore	len	Documentation
gb_est20:AA858296	-	98.00	1729.05	1.6e-87	609	AA858296 ob13b06.s1 NCI_CGAP_K1
gb_est24:AI822455	-	72.00	1271.07	5.3e-62	243	AI822455 q921b09.x1 NCI_CGAP_K1
gb_est34:AI922430	-	66.00	1163.87	4.9e-56	244	AI922430 q921b09.y5 NCI_CGAP_K1
gb_est24:AI733707	-	62.00	1091.06	5.6e-52	297	AI733707 q921b09.x5 NCI_CGAP_K1
gb_est33:AI821963	-	61.00	1075.58	4.1e-51	211	AI821963 q921b09.x5 NCI_CGAP_K1
gb_est18:AA706766	-	44.00	769.01	1.9e-34	131	AA706766 z120d11.x1 Soares_feta
gb_est18:AI247271	-	16.00	273.21	2.0e-06	173	AI247271 q160f11.x1 NCI_CGAP_K1
gb_gss3:BI73160	-	11.00	176.57	0.4862	455	BI73160 RPI11-11K11.PP RPI1-11
gb_est24:AI196449	-	10.00	159.31	4.45	454	AI196449 u169h08.x1 Sngano_mous
gb_est41:AV67452	-	9.00	145.07	27.60	270	AV67452 AV267452 RIKEN full-16
gb_est27:AI27938	-	9.00	141.48	43.78	452	AI27938 mm25d12.x1 Stratagene
gb_est12:AA276173	-	9.00	140.85	47.48	495	AA276173 vc31a07.x1 Barstead_MF
gb_gss15:AO570823	-	9.00	140.79	47.83	499	AO570823 HS-5366.B1.F11.SP6E.RF
gb_est38:AA067082	-	9.00	140.46	49.88	523	AA067082 683015C02.x1 683 - 14
gb_gss3:BI0499	-	9.00	135.36	95.92	1086	BI0499 T6A18-SP6.2.TAMU Arabid
gb_est1:TI2786	-	8.00	130.86	170.81	160	TI2786 0024 Pyrococcus furiosus
gb_est20:AA507561	-	8.00	130.52	178.43	168	AA507561 om09e01.x1 Soares_NFL
gb_est18:AA010571	-	8.00	127.75	254.66	250	AA010571 z109c09.s1 Soares_feta
gb_est9:DI76900	-	8.00	127.29	270.10	267	DI76900 MUSE6B04 mouse embryonal
gb_est4:HA2487	-	8.00	127.06	278.23	276	HA2487 Y141606.s1 Soares_breast
gb_est40:AV228411	-	8.00	126.89	284.54	283	AV228411 AV228411 RIKEN full-16
gb_est44:AA183544	-	8.00	126.69	291.73	291	AA183544 xj75g12.x1 Soares_NFL
gb_est1:Z46183	-	8.00	126.62	294.42	294	Z46183 HSC25C11 normalized inf
gb_est46:AA539493	-	8.00	126.60	295.31	295	AA539493 46118 MARC 2P1G Sus sc
gb_est4:HA3456	-	8.00	126.48	299.79	300	HA3456 EST111574 Rat PC-12 cell
gb_est21:DA2653	-	8.00	126.41	302.47	303	DA2653 DA2653 Rice callus cDNA
gb_gss1:CN500DMH	-	8.00	126.23	309.61	311	AI091175 Arabidopsis thaliana g
gb_gss3:BA4529	-	8.00	126.23	309.61	311	BA4529 HS-1059-B1-D02-MF.ab1.C1
gb_est9:AA090850	-	8.00	126.07	315.84	318	AA090850 yu0491.seq.F Human fet
gb_gss3:BA8504	-	8.00	125.98	319.39	322	BA8504 RPI11-11K11.PP RPI1-11 H
gb_est9:AA116763	-	8.00	125.56	337.08	342	AA116763 mq25e05.x1 Barstead_MF
gb_est35:AI825450	-	8.00	125.38	345.01	351	AI825450 wb73b01.x1 NCI_CGAP_K1
gb_gss3:BA04650	-	8.00	125.36	345.89	352	BA04650 cSRL-4069-u CSL flow sc
gb_est26:AI159583	-	8.00	125.17	354.67	362	AI159583 qy22d11.x1 NCI_CGAP_K1
gb_est23:AI606736	-	8.00	125.17	354.67	362	AI606736 vc15e05.y1 Barstead_MF
gb_est6:NA5217	-	8.00	125.11	357.30	365	NA5217 20257 Lambda-PRI2 Arabid
gb_gss10:AA0107667	-	8.00	125.07	359.05	367	AA0107667 HS-3111.AI-H09.MR.C1T
gb_est5:HA6779	-	8.00	125.03	360.80	369	HA6779 yr84g10.r1 Soares fetal

gb_est7:WA1381 - 8.00 125.00 362.55 371 WA1381 mc46e01.r1 Soares mou
 gb_est17:CG6800 - 8.00 124.92 366.05 375 CG6800 CG6800 Yuj1 Kohara un
 gb_gss11:AA0191156 - 8.00 124.90 366.92 376 AA0191156 HS-2195-B2_A05_MR C
 gb_est9:AA104994 + 8.00 124.88 367.79 377 AA104994 m180c05.x1 Stratage
 gb_est29:AI587324 + 8.00 124.74 374.77 385 AI587324 tq04h05.x1 NCI_CGAP
 gb_est46:AA326686 + 8.00 124.74 374.77 385 AA326686 19580 MARC 2B0V Bos

seq_name: gb_est20:AA858296

seq_documentation_block:

LOCUS AA858296 609 bp mRNA 09-MAR-1998
 DEFINITION ob13b06.s1 NCI_CGAP_K1d3 Homo sapiens cDNA clone IMAGE:1323539 3'
 similar to TR:0004476 0004476 SODIUM PHOSPHATE TRANSPORTER. ;, mRNA
 sequence.

ACCESSION AA858296
 VERSION AA858296.1 GI:2946598
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 609)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Jan 19, 1998 this sequence version replaced gi:2285607.
 CONTACT Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN at:
 www.bio.llnl.gov/dbfp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 467.
 Location/Qualifiers

FEATURES
 source 1..609
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1323539"
 /clone_lib="NCI_CGAP_K1d3"
 /lab_host="DH10B"

/note="organ: kidney; Vector: p773D-Pac (pharmacia) with
 a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - Oligo(dT) primer,
 double-stranded cDNA was ligated to Eco RI adaptors
 (pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified p773 vector. mRNA
 source: 2 pooled kidneys. Library went through one round
 of normalization. Library constructed by Bento Soares and
 M. Fatima Bonaldo."

BASE COUNT 166 a 118 c 114 g 211 t

ORIGIN
 alignment_scores: 98.00 Length: 98
 Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-391-958-1 x AA858296/rev ..

Align seq 1/1 to reverse of: AA858296 from: 1 to: 609

304 Thrleusercysglleuserthrleucysglserlylerylrleas 320
 ||||||||||||||||||||||||||||||||||||||||||

603 ACGCTCTTCGGATTAAGCACATTGTGTCAGTCAGGCAATTATATCA 554
 320 nValleuaspIleAlaProArGTYrSerSerPheLeuMetGlyAlaSerA 337
 |||||||
 553 TGTCTTAGATATGTCTCAAGGATATTCAGATTTCTCATGGAGACATCA 504
 337 rGgIyPheSerSerIleAlaProValIleValIProThrValSerGlyPhe 353
 |||||||
 503 GAGGATTTTCAGCATAGCACACCTGTCATTTGACCCACTGTCAGCGGATTT 454
 354 LeuLeuSerGlnAspProGluPheGlyTPrParGsnValPhePheLeu 370
 |||||||
 453 CTCTTAGTCAGGACCTGAGTTGGGTGAGAGATCTCTTCTTCTTCT 404
 370 uPheAlaValAsnLeuLeuGlyLeuLeuPheTYrLeuIlePheGlyGlu 387
 |||||||
 403 GTTTCGCGTACCTGTTAGGACTACTCTTCACTCATATTTGGAGAAC 354
 387 lAspValGlnGluTPrAlaValGlnArgLysLeuThrArgLeu 401
 |||||||
 353 CAGATGTCCAGAAATGGCTTAAGAGAAACCTCACTCGTTTA 310

seq_name: gb_est24:AI222455

seq_documentation_block: 243 bp mRNA

LOCUS AI222455 27-OCT-1998
 DEFINITION qg21b09.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1761689 3',
 similar to TR:000476 000476 SODIUM PHOSPHATE TRANSPORTER. ;, mRNA
 sequence.
 AI222455
 VERSION AI222455.1 GI:3804658
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On May 9, 1995 this sequence version replaced gi:803109.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.lnl.gov/bdrrp/image/image.html

Seq primer: -400P from Gibco
 High quality sequence stop: 164.

FEATURES

source

1. 243

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1761689"

/clone_1lb="NCI_CGAP_Kid3"

/lab_host="DH10B"

/note="Organ: Kidney; Vector: p773D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dt) primer,

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified p773 vector. mRNA

source: 2 pooled kidneys. Library went through one round

of normalization. Library constructed by Bento Soares and

M. Fatima Bonaldo."

BASE COUNT

75 a 57 c 50 g 61 t

ORIGIN

alignment_scores: 72.00 Length: 72
 Quality: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x AI222455/rev ..

Align seg 1/1 to reverse of: AI222455 from: 1 to: 243

330 SerPheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaProValI 346
 |||||||
 243 AGTTTCTCATGGAGCATCAAGATTTTTCAGCATATACCTGATC 194
 346 eValProThrValSerGlyPheLeuLeuSerGlnAspProGluPheGly 363
 |||||||
 193 TGTACCCACTGTCAGCGGATTTCTTCTTAGTCAAGACCTGAGTTGGGT 144
 363 rParGsnValPhePheLeuPheAlaValAsnLeuLeuGlyLeuLeu 379
 |||||||
 143 GGAGGAATGCTCTTCTTCTGCTGTTGCCGTAACTGTTAGGACTACTC 94
 380 PheTYrLeuIlePheGlyGluAlaAspValGlnGluTPrAlaLysGlu 396
 |||||||
 93 TTCTACCTCATATTTTGGAGAACGATGTCCAGAAATGGCTTAAGAGAG 44
 396 gLysLeuThrArgLeu 401
 |||||||
 43 AAAAATCACTCGTTTA 28

seq_name: gb_est34:AI792430

seq_documentation_block: 244 bp mRNA

LOCUS AI792430 02-JUL-1999
 DEFINITION qg21b09.y5 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1761689 5',
 similar to SW:NP74_HUMAN 000476 SODIUM-DEPENDENT PHOSPHATE
 TRANSPORT PROTEIN 4 ;, mRNA sequence.
 AI792430
 VERSION AI792430.1 GI:5340146
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Dental Research,
 Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 On Jun 5, 1998 this sequence version replaced gi:3187571.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.lnl.gov/bdrrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 original clone citation: see original entry for original citation
 information
 This 5' resequenced clone has no previous 5' data to verify this
 new read against
 Putative full length read
 The vector to vector length is 245
 Seq primer: -400P from Gibco.

FEATURES
source Location/Qualifiers
1.244
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 61 a 49 c 55 g 79 t

ORIGIN

alignment_scores:
Quality: 66.00 Length: 66
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AT792430 ..

Align seg 1/1 to: AT792430 from: 1 to: 244

336 SerATGGLPheSerSerIleAlaProValIleValProThrValSerGI 352
|||||
19 TCAGAGAGATTTCGACGATAGCACCTCTCATTTGACCCATGTCAGCG 68
352 yPheLeuLeuSerGIAspProGIuPheGIyTrpArgAsnValPhePheL 369
|||||
69 ATTCTCTTAGTCGAGACCTGAGATTGGGTGGAGAAATGCTCTTCT 118
369 eulPheAlaValAsnLeuLeuGIyLeuLeuPheTyTrpLeuIlePheGI 385
|||||
119 TGGCTTTGCCGTTAACTGTTAGGACATACCTCTTACCTCATATTTCGA 168
386 GluAlaAspValGIuGIuTrpAlaLysGIuArgLysLeuThrArgLeu 401
|||||
169 GAACGAGATGTCGAAGATGGGCTAAAGAGAGAAACTCATTCTTTA 216

seq_name: gb_est24:AI244777

seq_documentation_block:
LOCUS AI244777 297 bp mRNA EST 28-JAN-1999
DEFINITION g192f01.x1 NCI-CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1866937 3'
similar to SW:NP74_HUMAN O00476 SODIUM-DEPENDENT PHOSPHATE
TRANSPORT PROTEIN 4 ; mRNA sequence.
ACCESSION AI244777
VERSION AI244777.1 GI:3840174
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 297)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced g1:2045043.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

JOURNAL
COMMENT

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.liml.gov/bdrp/image/image.html

Insert Length: 707 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 156.

FEATURES
source Location/Qualifiers
1.297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1866937"
/clone_lib="NCI-CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 94 a 68 c 59 g 75 t 1 others

ORIGIN

alignment_scores:
Quality: 62.00 Length: 62
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AI244777/rev ..

Align seg 1/1 to reverse of: AI244777 from: 1 to: 297

340 SerSerIleAlaProValIleValProThrValSerGIyPheLeuLeuSe 356
|||||
213 TCAGACCATGACACCTGATTTGATGACCATGACGATTTCTTCTAG 164
356 rGIAspProGIuPheGIyTrpArgAsnValPhePheLeuPheAlaV 373
|||||
163 TCAGACCATGATGTTGGGTGGAGAAATGCTCTTCTTCTGTTGCCG 114
373 aLAsnLeuLeuGIyLeuLeuPheTyTrpLeuIlePheGIyGluAlaAspVal 389
|||||
113 TTAACCTGTTAGGACTCTTCTTACCTCATTTTGGAGAACACATGTC 64
390 GIuGIuTrpAlaLysGIuArgLysLeuThrArgLeu 401
|||||
63 CAAGAAATGGGCTAAAGAGAGAAACTCATTCTTTA 28

seq_name: gb_est32:AT733707

seq_documentation_block:
LOCUS AT733707 211 bp mRNA EST 14-JUN-1999
DEFINITION g921b09.x5 NCI-CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1761689 3'
similar to SW:NP74_HUMAN O00476 SODIUM-DEPENDENT PHOSPHATE
TRANSPORT PROTEIN 4 ; mRNA sequence.
ACCESSION AT733707
VERSION AT733707.1 GI:5054820
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 211)
NCI/MIDR-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)

JOURNAL

COMMENT On Dec 20, 1995 this sequence version replaced gi:1133358.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbfp/image/image.html

seq_name: gb_est35:AI821963

seq_documentation_block:
LOCUS AI821963 452 bp mRNA EST 09-JUL-1999
DEFINITION ob13b06.x5 NCI CGAP Kid3 homo sapiens CDNA clone IMAGE:1323539 3'
TRANSPORT PROTEIN 4 ;, mRNA sequence.
ACCESSION AI821963

FEATURES
Source
Location/Qualifiers
1..211
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1761689"
/clone_lib="NCI-CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

BASE COUNT 64 a 49 c 45 g 53 t

ORIGIN

alignment_scores:
Quality: 61.00 Length: 61
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AI733707/rev ..

Align seg 1/1 to reverse of: AI733707 from: 1 to: 211

341 SerTlleAlaProValIleValProThrValSerGlyPheLeuSerGI 357
|||||
210 AGCATAGCACCTGTCATGTACCCACTGCAGCGATTCTCTTAGTCA 161
|||||
357 naspprogiluphegilytrpargasnValphepheleuleuphealVala 374
|||||
160 GAACCTGAGTTGGGTGGAGAGATGTCCTTCTGCTGCGGTTA 111
|||||
374 snleuleuglyleuleupheyleuilephegilyGluAlaaspValgin 390
|||||
110 ACCGTGTAGACTACTCTCTACCTCATATTGAGAGAACAGATGCCAA 61
|||||
391 GluTTPAlaLysGluArgLysLeuThrArgLeu 401
|||||
60 GAATGGGCTAAAGAGAGAAACTCACTCTTAA 28
|||||

VERSION AI821963.1 GI:5441042
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 452)
NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute / National Institute of Dental Research,
TITLE Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
On Mar 10, 1998 this sequence version replaced gi:2948996.
Other ESTs: ob13b06.y5
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbfp/image/image.html

seq_name: gb_est35:AI821963

seq_documentation_block:
LOCUS AI821963 452 bp mRNA EST 09-JUL-1999
DEFINITION ob13b06.x5 NCI CGAP Kid3 homo sapiens CDNA clone IMAGE:1323539 3'
TRANSPORT PROTEIN 4 ;, mRNA sequence.
ACCESSION AI821963

FEATURES
Source
Location/Qualifiers
1..452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1323539"
/clone_lib="NCI-CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

BASE COUNT 115 a 84 c 78 g 175 t

ORIGIN

alignment_scores:
Quality: 47.00 Length: 47
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AI821963/rev ..

Align seg 1/1 to reverse of: AI821963 from: 1 to: 452

355 LeuSerGlnaspProGiluphegilytrpargasnValphepheleuleuph 371
|||||
450 CTAGTCAGAGACCCCTGAGATTGGGTGGAGAGATGTCCTTCTGCTGTT 401
|||||
371 eAlaValaAsnleuleuglyleuleupheyleuilephegilyGluAla 388
|||||
400 TGGCGTTAAGCTGTTAGACTACTCTTACCTCATATTGAGAGAGCG 351
|||||
388 spValGlnGluTTPAlaLysGluArgLysLeuThrArgLeu 401
|||||
350 ATGTCCAGAAATGGGCTAAAGAGAGAAACTCACTCTGTTA 310
|||||

```

|||||
166 TAGTACGAGCCTTGACTTTGGTGGAGAGATCTCTCTCTGTTGTTTG 117
|||||
372 JavalAsnLeuEngLyLeuLeuPheTyRLeu 382
|||||
116 CCGTTAACCTGTTAGGACTACTCTCTACCTT 85
|||||
seq_name: gb_est25:AI247271

seq_documentation_block:
LOCUS      173 bp      mRNA      EST      28-JAN-1999
DEFINITION GI604771 x1 NCI CGAP Kid3 Homo sapiens cDNA clone IMAGE:1863885 3'
            similar to SW:NEPT4_HUMAN O00476 SODIUM-DEPENDENT PHOSPHATE
            TRANSPORT PROTEIN 4 ;, mRNA sequence.
ACCESSION  AI247271
VERSION    AI247271.1 GI:3842668
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 173)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
TITLE       Unpublished (1997)
            On Jan 19, 1998 this sequence version replaced gi:228652.
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/BLM at:
            www-bio.litl.gov/bdrip/image/image.html

Insert Length: 1365 Std Error: 0.00
Seq primer: -40UP from GIBCO
High quality sequence stop: 125.
Location/Qualifiers
1. .173
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="IMAGE:1863885"
   /clone_lib="NCI-CGAP_Kid3"
   /lab_host="DH10B"
   /note="Organ: kidney; Vector: pTV73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pTV73 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo.
BASE COUNT  52 a      41 c      34 g      45 t      1 others
ORIGIN
alignment_scores:
Quality:      16.00      Length:      16
Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-391-958-1 x AI247271/rev ...
Align seg 1/1 to reverse of: AI247271 from: 1 to: 173
385 GtGtGtAaTaaPvaIcInGtUtrPalatYsgUatRgLyIsUeuthrAag 400

```


124 a	129 c	88 g	154 t
-------	-------	------	-------

alignment_block:

US-09-391-958-1 x A0570823 ..

Align seg 1/1 to: A0570823 from: 1 to: 499

302 LeuLeuThrLeuSerCysGlyLeuSer 310
|||||
458 CTGTTGACTCTCTCGTGTGGTTATCC 484

seq_name: gb_est38:AW067082

seq_documentation_block:

LOCUS AW067082 523 bp mRNA EST 12-OCT-1999
DEFINITION 683015C02.x1 683 - 14 day immature embryo from Hake lab (HS) Zea
mays cDNA, mRNA sequence.

ACCESSION AW067082
VERSION AW067082.1 GI:6022154

KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Zea.
1 (bases 1 to 523)

REFERENCE
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished (1999)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135638.
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 683015 row: C column: 02.

FEATURES
source 1..523
Location/Qualifiers
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="683 - 14 day immature embryo from Hake lab
(HS)"
/tissue_type="embryo"
/dev_stage="14 days after pollination"
/lab_host="DH10B"
/note="Organ: embryo; Vector: PBKCMV (Stratagene's Zap
Express); Site_1: XhoI; Site_2: EcoRI; directionally
cloned, 14 day immature embryo library created with
Stratagene's Zap Express cDNA protocol."

BASE COUNT 92 a 160 c 177 g 94 t
ORIGIN

alignment_scores:

Quality:	9.00	Length:	9
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-391-958-1 x AW067082 ..
Align seg 1/1 to: AW067082 from: 1 to: 523

277 LeuGlySerLeuProSerSerAlaLeu 285
|||||
78 CTCGATCTCTCTCCCTCCGCTTA 104

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 13:32:03 ; Search time 28.66 Seconds

(without alignments)
331.407 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 2076
Sequence: 1 MOVDETLIPRKVPISCSARY.....LIFGEADVQIMAKERKUTRL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2076	100.0	401	1 W69971	Human sodium-depen
2	2061	99.3	401	1 W78920	Human haemochromat
3	744	35.8	436	1 W78919	Human haemochromat
4	479	23.1	560	1 W05148	Human brain sodium
5	479	23.1	560	1 W70500	Human sodium-lithi
6	466.5	22.5	576	1 W88523	Eat-4 protein amin
7	169.5	8.2	482	1 W14439	Protein involved i
8	122.5	5.9	462	1 W97820	Staphylococcus aur
9	112.5	5.4	2969	1 W56446	Fragment HGJ2141 o
10	111.5	5.4	2965	1 W56450	Fragment F1029 of
11	111.5	5.4	2970	1 W56445	Fragment HGJ1916 o
12	110.5	5.3	2873	1 W56441	Fragment HGJ606 of
13	110.5	5.3	2969	1 W56443	Fragment HGJ1741 o
14	108.5	5.2	3011	1 R95020	Hepatitis GB virus
15	107.5	5.2	470	1 R99353	Human amine transp
16	107.5	5.2	2860	1 W63611	Hepatitis GB virus
17	107.5	5.2	2905	1 W06536	Hepatitis GB virus
18	107.5	5.2	2969	1 W56442	Fragment HGJ1737 o
19	106.5	5.1	2873	1 R87559	Hepatitis virus cl
20	106.5	5.1	2873	1 R90796	HGV-PNF 2161 polyp
21	106.5	5.1	2873	1 W34983	Hepatitis G virus
22	106.5	5.1	2873	1 W76054	HGV isolate PNF 21
23	106.5	5.1	2873	1 W80148	HGV-PNF2161 varian
24	106.5	5.1	2873	1 W89452	Hepatitis G virus
25	106.5	5.1	2873	1 W87355	US5856134 Seq ID 1
26	106.5	5.1	2969	1 W56449	Fragment K3732(2v)
27	104.5	5.0	470	1 W77493	Human amine transp
28	104.5	5.0	2910	1 R87566	Hepatitis G virus
29	104.5	5.0	2910	1 R80797	Hepatitis G virus
30	104.5	5.0	2910	1 W76088	HGV-JC variant pol
31	104.5	5.0	2910	1 W80182	Hepatitis G virus
32	104.5	5.0	2910	1 W89458	Hepatitis G virus
33	104.5	5.0	2910	1 W92789	US5856134 Seq ID 1
34	104	5.0	502	1 W99600	B.stations mult-

35	103.5	5.0	488	1 W37990	Mutant Aspergillus
36	103.5	5.0	514	1 R47342	Mammalian synaptic
37	103.5	5.0	514	1 W38286	Human synaptic ves
38	103.5	5.0	561	1 W14440	Protein Involved i
39	103.5	5.0	2723	1 W56448	Fragment HGJ1775 o
40	103	5.0	606	1 W35808	Cercospora kikuchi
41	102.5	4.9	559	1 W64554	Human liver cell c
42	102.5	4.9	2841	1 W06537	Hepatitis GB virus
43	101.5	4.9	537	1 W41195	Mouse osteoclast t
44	99.5	4.8	2969	1 W56447	Fragment HGJ168 o
45	99	4.8	456	1 P91895	Protein sequence o

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description
1	W69971	100.0%	100.0%	401	1 W69971	Human sodium-depen
2	W69971	99.3%	100.0%	401	1 W78920	Human haemochromat
3	W69971	35.8%	100.0%	436	1 W78919	Human haemochromat
4	W69971	23.1%	100.0%	560	1 W05148	Human brain sodium
5	W69971	23.1%	100.0%	560	1 W70500	Human sodium-lithi
6	W69971	22.5%	100.0%	576	1 W88523	Eat-4 protein amin
7	W69971	8.2%	100.0%	482	1 W14439	Protein involved i
8	W69971	5.9%	100.0%	462	1 W97820	Staphylococcus aur
9	W69971	5.4%	100.0%	2969	1 W56446	Fragment HGJ2141 o
10	W69971	5.4%	100.0%	2965	1 W56450	Fragment F1029 of
11	W69971	5.4%	100.0%	2970	1 W56445	Fragment HGJ1916 o
12	W69971	5.3%	100.0%	2873	1 W56441	Fragment HGJ606 of
13	W69971	5.3%	100.0%	2969	1 W56443	Fragment HGJ1741 o
14	W69971	5.2%	100.0%	3011	1 R95020	Hepatitis GB virus
15	W69971	5.2%	100.0%	470	1 R99353	Human amine transp
16	W69971	5.2%	100.0%	2860	1 W63611	Hepatitis GB virus
17	W69971	5.2%	100.0%	2905	1 W06536	Hepatitis GB virus
18	W69971	5.2%	100.0%	2969	1 W56442	Fragment HGJ1737 o
19	W69971	5.1%	100.0%	2873	1 R87559	Hepatitis virus cl
20	W69971	5.1%	100.0%	2873	1 R90796	HGV-PNF 2161 polyp
21	W69971	5.1%	100.0%	2873	1 W34983	Hepatitis G virus
22	W69971	5.1%	100.0%	2873	1 W76054	HGV isolate PNF 21
23	W69971	5.1%	100.0%	2873	1 W80148	HGV-PNF2161 varian
24	W69971	5.1%	100.0%	2873	1 W89452	Hepatitis G virus
25	W69971	5.1%	100.0%	2873	1 W87355	US5856134 Seq ID 1
26	W69971	5.1%	100.0%	2969	1 W56449	Fragment K3732(2v)
27	W69971	5.0%	100.0%	470	1 W77493	Human amine transp
28	W69971	5.0%	100.0%	2910	1 R87566	Hepatitis G virus
29	W69971	5.0%	100.0%	2910	1 R80797	Hepatitis G virus
30	W69971	5.0%	100.0%	2910	1 W76088	HGV-JC variant pol
31	W69971	5.0%	100.0%	2910	1 W80182	Hepatitis G virus
32	W69971	5.0%	100.0%	2910	1 W89458	Hepatitis G virus
33	W69971	5.0%	100.0%	2910	1 W92789	US5856134 Seq ID 1
34	W69971	5.0%	100.0%	502	1 W99600	B.stations mult-

```

OY 181 VSSKQPLPIKAMLSPLWISICLCFHSQWLVSTMVYIPIYISSYHVNRDNGLLSA 240
    |||||||
DB 181 VSSKQPLPIKAMLSPLWISICLCFHSQWLVSTMVYIPIYISSYHVNRDNGLLSA 240
OY 241 LPIYAWYIGWGVGLADFLTLTKRRLTVRKRIATILGSLPSALIVSLPYLNSGYTAT 300
    |||||||
DB 241 LPIYAWYIGWGVGLADFLTLTKRRLTVRKRIATILGSLPSALIVSLPYLNSGYTAT 300
OY 301 ALLTISGISTLCOSGIYINVDIAPRYSSFLMGASRGSSIAPIYPTVSGFLLSQDE 360
    |||||||
DB 301 ALLTISGISTLCOSGIYINVDIAPRYSSFLMGASRGSSIAPIYPTVSGFLLSQDE 360
OY 361 FGMNVFFLLFAVNLGLLFYLIIFGEADVQEWAKERKRLRL 401
    |||||||
DB 361 FGMNVFFLLFAVNLGLLFYLIIFGEADVQEWAKERKRLRL 401

RESULT 2
W78920
ID W78920 standard; Protein; 401 AA.
AC W78920:
DE 21-DEC-1998 (first entry)
KW Human haemochromatosis protein NRP4.
KW Bovine butyrophilin; BR; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BRP1; BRP2; BRP3;
KW BRP4; BRP5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene.
OS Homo sapiens.
PN MO9814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
PI WPI: 98-240014/21.
DR N-PSDB: V57910.
PT Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 60; Fig 5B: 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BR genes, which are homologues of the milk
CC protein butyrophilin (BR), and can be used in the production of agonists
CC and antagonists of BR function. Also described are: (1) a Roret gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia. The present sequence represents
CC NRP4.
SQ Sequence 401 AA:

Query Match 99.3%; Score 2061; DB 1; Length 401;
Best Local Similarity 99.3%; Pred. No. 1.3e-211;
Matches 396; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

DB 61 SEVLVDSFGSLGSKAPKSLPAKSSILGQFAIWERWGPQERSRLCSIALSGMLGCFTA 120
    |||||||
OY 121 ILIGFISFETLGMPFVFYIFGGVCCCLMPFVYIDDPVSYPWISTEKEKYEIISLAKQ 180
    |||||||
DB 121 ILIGFISFETLGMPFVFYIFGGVCCCLMPFVYIDDPVSYPWISTEKEKYEIISLAKQ 180
OY 181 VSSKQPLPIKAMLSPLWISICLCFHSQWLVSTMVYIPIYISSYHVNRDNGLLSA 240
    |||||||
DB 181 VSSKQPLPIKAMLSPLWISICLCFHSQWLVSTMVYIPIYISSYHVNRDNGLLSA 240
OY 241 LPIYAWYIGWGVGLADFLTLTKRRLTVRKRIATILGSLPSALIVSLPYLNSGYTAT 300
    |||||||
DB 241 LPIYAWYIGWGVGLADFLTLTKRRLTVRKRIATILGSLPSALIVSLPYLNSGYTAT 300
OY 301 ALLTISGISTLCOSGIYINVDIAPRYSSFLMGASRGSSIAPIYPTVSGFLLSQDE 360
    |||||||
DB 301 ALLTISGISTLCOSGIYINVDIAPRYSSFLMGASRGSSIAPIYPTVSGFLLSQDE 360
OY 361 FGMNVFFLLFAVNLGLLFYLIIFGEADVQEWAKERKRLRL 401
    |||||||
DB 361 FGMNVFFLLFAVNLGLLFYLIIFGEADVQEWAKERKRLRL 401

RESULT 3
W78919
ID W78919 standard; Protein; 436 AA.
AC W78919:
DE 21-DEC-1998 (first entry)
KW Human haemochromatosis protein NRP3.
KW Bovine butyrophilin; BR; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BRP1; BRP2; BRP3;
KW BRP4; BRP5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene.
OS Homo sapiens.
PN MO9814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
PI WPI: 98-240014/21.
DR N-PSDB: V57909.
PT Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 54; Fig 5B: 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BR genes, which are homologues of the milk
CC protein butyrophilin (BR), and can be used in the production of agonists
CC and antagonists of BR function. Also described are: (1) a Roret gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia. The present sequence represents
CC NRP3.
SQ Sequence 436 AA:

Query Match 35.8%; Score 744; DB 1; Length 436;
Best Local Similarity 38.5%; Pred. No. 4.5e-71;
Matches 162; Conservative 56; Mismatches 107; Indels 96; Gaps 5;

```

Page 3

```

OY 72 LSKAKSLPAMSSILGGGFALMERNGPQDESRSLCSLALSMGLGCFALIGFISETL 133
Db 191 -----HGIMSKAPPLESLRLATTAFCGSYAGAVAMPALVLVOYS 233
OY 132 GMPFVFIYFGGVCACCLMFVEVIYIDDPVSXPWISTSEKEYIISLKOQVSSKO---PL 188
Db 233 GMSVFYIYGSIGITWYLFWLLVYESPALPISISEEKRI-----EADIGESKLMNPL 288
OY 189 -----PIKAMLRSLPWSICLACGCSHOWLSTVMVVIPTIYISSYHVNIIRDNLALPF 243
Db 289 TKFSTPMRRFFFSMPVVAIIVANFCRSWTFYLLLSIDPDYEEFVEFGFELISVGLVALPH 348
OY 244 IVANWIGWVGGLADELLTKKFRLLT-VRKATFIIGLSPSALLVSLPYLNSGITATLAL 302
Db 349 LVMITVILPIGGIADPLSKRIMSTTNVTKMLKMGCGEGMEATLLLVGVSHSKSV-AISF 407
OY 303 LTLSCGLSTLCSGCIYINVLDIAPRSSFLGASGKGFSSIPAVIYVPGCLFSQDEPFG 366
Db 408 LVLANGFSGFPAISGNVNHDLIAPRYAAILMGISNGVGTLSGMCPITVGMATHKTRRE 467
OY 363 WRNVEFLFAVNLGLLFYLLIFGEADVOEMAKERKLT 399
Db 468 WQYVFLIASLVHGGVIFGYGFASEKGPMAPEBMS 504

RESULT 5
W70500
AC W70500 standard; Protein; 560 AA.
DT 18-JAN-1999 (first entry)
DE Human sodium-lithium countertransporter BNPI.
KM Sodium-lithium countertransporter; sodium phosphate cotransporter;
KM BNPI; human; lithium therapy; manic depression.
OS Homo sapiens.
PN W093838203-A1.
PD 03-SEP-1998
PF 11-FEB-1998; 002875.
PI 27-FEB-1997; US-039462.
PR (UYEM-) UNIV EMORY.
PI Gunn RB, Timmer RT;
DR WPI; 98-520759/44.
DR N-PSDB; V33503.
PT New isolated lithium-sodium counter-transporter DNA - used to
PT develop products for evaluating lithium-sodium transport in
PT erythrocytes, particularly for lithium therapy in manic depression.
PS Claim 8; Page 45-47; 64pp; English.
CC This polypeptide comprises the brain-specific human lithium-sodium
CC countertransporter (LSCT) BNPI. A DNA sequence for human BNPI is
CC provided (see V33503). LSCTs such as BNPI provide the
CC physiological mechanism for the extrusion of lithium from cells,
CC i.e. it regulates the cell concentration of lithium. Its activity
CC determines the therapeutic effect of lithium. The invention
CC provides a simple molecular biological test for the ability of
CC cells to extrude lithium. The LSCTs have significance for
CC determining the responsiveness of humans with mental disorders,
CC including manic depressives, to treatment with lithium salts.
CC Probes and primers for BNPI, PIT-1 (see W70498), PIT-2 (see
CC W70499) can be used in diagnostic tests useful for genetic
CC screenings to predict whether a patient will respond to lithium
CC treatment. The test is also a screen for susceptibility to, and
CC extent of, manic depressive illness, and is suitable for screening
CC newborns.
SQ Sequence 560 AA;

```

Query Match	23.1%	Score 479;	DB 1;	Length 560;
Best Local Similarity	28.2%	Pred. No. 1.2e-42;		
Matches 112: Conservative	68;	Mismatches 155;	Indels 62;	Gaps 7;
QY	12	VSLSGARIGIALVLFHCFNFTTIAOVINNTIMVAMVNSTPQSLINDSEVLVPDSEFG	71	
		: :		
DB	161	ISASARVHGVCYI-----FVRILGIVLEGVCYIPPC-----	191	
		: :		

QY 72 LSKAPKSLPANSSILGGFAIERNWGPPEERSRLCSIALSGMLGCTFAILLIGFISETL 131
DB 191 -----HGIIWSKMAPPLERSRLATAFCCSGYAGAVVAMPALVLYOS 232
QY 132 GMPFEYFIYFGVGCYCCLLMFVYIYDDDEVSYPMWISTSEKEYIISLKQOVGSKQ---PL 188
DB 233 GMSVFYVYGSFGIFWYIFWLLVSYESPALHPSISEERKTY----EADIGESALIMPL 288
QY 189 -----PIKAMLRSLPIWISICLCCFSHOWLVSTMVYIPIYISSYHVNRDGLSALPE 243
DB 289 TKFSIPWRFSTSMRYVALIVANFCRSMWTFYILLISQPDYEEVGFELSKYGLVSALPH 348
QY 244 IVANVIGWGYLADFLTKRFLIT-VKRIATILGSLPSALIVSLPYNSGYITATL 302
DB 349 LVMITIVIGQIADFLRSRIMSTTVNRKILMNCGGFGEATLLLVGYSKGV-AISF 407
QY 303 LTLSCGLTLCOSGIYINVLDAIPRYSFIMKASRGFSSIAVYIPYVSGFLISDPERG 362
DB 408 LVLAVFGFALISGFNVHLDIAPRYASIIIMGINSVGTLSGMVCPITIVGAMTKHKTRBE 467
QY 363 WRNVFFLLFAVNLGLLFYILFGEADVQEMAKERKIT 399
DB 468 WQYVPLIASIVHYGVITFYGVFASGEKQPMAPPEMS 504

RESULT 6
W88523
ID W88523 standard; protein; 576 AA.
AC W88523:
DT 05-MAR-1999 (first entry)
DE Eat-4 protein amino acid sequence.
KW Eat-4 protein; CAMP regulatory gene; neurodegeneration; CAMP;
KW cyclic adenosine monophosphate; reporter gene; Alzheimer's; Parkinson's;
KW Huntington's disease; amyotrophic lateral sclerosis; multiple sclerosis;
KW stroke; epilepsy; GTP-binding protein G alpha s; Gas;
KW adenyllyl cyclase gene; Gas-induced cell death; CAMP level.
OS Caenorhabditis elegans.
PN W09853856-A1.
PD 03-DEC-1998.
PE 29-MAY-1998; U11058.
PR 29-MAY-1997; US-864785.
PA (GENO) GEN HOSPITAL CORP.
PI Hart AC, Kaplan JM, Oppenheimer AJ;
DR WPI; 99-045345/04.
PT Identifying agents for treating or preventing neurodegeneration -
PT by assessing their effect on CAMP regulatory genes or promoters.
PT useful for, e.g. potentially treating Alzheimer's disease
PS Disclosure; Fig 3; 66pp; English.
CC The present sequence represents an eat-4 protein of Caenorhabditis
CC elegans. The eat-4 gene is a CAMP regulatory gene. The specification
CC describes the identification of compounds for treating (or preventing
CC onset of) neurodegeneration. The method comprises treating a cell having
CC a reporter gene linked to a CAMP (cyclic adenosine monophosphate)
CC regulatory gene or promoter, with a test compound and measuring
CC expression of the reporter gene. Such compounds are potentially useful
CC for treating and preventing Alzheimer's, Parkinson's and Huntington's
CC diseases, amyotrophic lateral sclerosis, multiple sclerosis, stroke or
CC epilepsy. The methods are based on the observation that, in the nematode
CC Caenorhabditis elegans, constitutive activation of the GTP-binding
CC protein G alpha s (Gas) induces neurodegeneration, and that mutations in
CC some genes, including adenyllyl cyclase-type genes, block Gas-induced cell
CC death, suggesting that the effect of Gas is associated with changes in
CC CAMP levels.
SQ Sequence 576 AA;

Query Match 22.5%; Score 466.5; DB 1; Length 576;
Best Local Similarity 32.3%; Pred. NO. 2,7e-41;
Matches 101; Conservative 65; Mismatches 134; Indels 13; Gaps 6;
QY 92 IWRNMGPPQERSRLCSIALSGMLGCTFAILLIGFISETLMPVYFYIFGVGCYCCLLM 151
DB 198 WMRVWAPMERSKLATWATGTSYAGAVLGLPLSAFLVSYVSWMAAFYLYGVGYVIMILM 257

QY 152 FVYIYDDPVSYPMWISTSEKEYIISLKQOVG---SSKOP-----LPKAMLRSLPIWISICLG 205
DB 258 FCVTFEKAFAHPTISOEKIFL-----EADIGHVSTHTPTINSIPKALVTSKPVAILVA 313
QY 206 CFSHOWLVSTMVYIPIYISSYHVNRDGLSALPEITVAVMVGWGYLADFLTKRKF 265
DB 314 NFARSMTFYLLQNLQTFYWKREALKGIADSGLLAIIPHLVNGCVVLMGGQADLYRSNKI 373
QY 266 RLIT-VKRIATILGSLPSALIVSLPYNSGYITATLITLSCGLTLCOSGIYINVIDI 324
DB 374 LSTTAVRKIFMCGGFGGAAFMILVAYTTSD-TTAIMLILAVGSGFASGFVNHLDI 432
QY 325 APRYSFIMKASRGFSSIAVYIPVSGFLISQDEFGMRNVFLLFAVNLGLFYILF 384
DB 433 APRYAIIMGSGNSGICITAGLTCFPVYTAFTAHK-KHGMTSVFLASLIHFTGYTFYAVY 491
QY 385 GEADVQEMAKERK 397
DB 492 ASGELQEWNAEPKE 504

RESULT 7
W14439
ID W14439 standard; Protein; 482 AA.
AC W14439:
DT 15-MAY-1997 (first entry)
DE Protein involved in cephalosporin C biosynthesis.
KW cephalosporin C; biosynthesis; fermentation.
OS Acromonium chrysogenum.
PN J09009966-A.
PD 14-JAN-1997.
PE 03-JUN-1995; 167461.
PR 03-JUL-1995; JP-167461.
PA (ASAH) ASAH KASEI KOGYO KK.
DR WPI; 97-126424/12.
DR N-PSDB; T62652-53.
PT Acromonium chrysogenum gene involved in biosynthesis of
PT cephalosporin C - used to improve fermentation ability of A.
PT chrysogenum
PS Claim 2; Page 9-11; 21pp; Japanese.
CC This protein, encoded by DNA derived from Acromonium chrysogenum, is
CC involved in biosynthesis of cephalosporin C. The gene involved in
CC biosynthesis of cephalosporin C or its cDNA can be used to improve the
CC fermentation ability of Acromonium chrysogenum.
SQ Sequence 482 AA;

Query Match 8.2%; Score 169.5; DB 1; Length 482;
Best Local Similarity 24.2%; Pred. NO. 9.9e-10;
Matches 81; Conservative 62; Mismatches 117; Indels 75; Gaps 17;
QY 100 QERSRLCSIALSGMLGCTFAILLIGFISETLGMPEFYIFGVGCYCCLLMFVYIYDDP 159
DB 165 EERGAAMAIWFGPLMGAVPAVAGCYLALEAKGMWVWVAIIGGFTTGAFILARF-- 223
QY 160 VSYPIISSEKEYIISLKQOVGSSKQPL-----PIKAMLRSLP 198
DB 223 -TTPVLLQKR---VNRKQETGN---PLTSALADISSRRARISRSYRPLVLLFRS-P 274
QY 199 I---WSICIG-CFSHOWLVSTMVYIPIYISSYHVNRDGLSALPEITVAVYWG-VWG 253
DB 275 IVPLFVSFIVAVFSYQFL---LFTYIPVFEIYDFSLGQIG-LSYLGIAAGLLGNATIF 330
QY 254 GYLADFLTKR-----RLITVKRIATILGSLPSALIVSLPYNSGYIT----- 299
DB 331 GOASDRILSKKSGMEKLEPERLPLM-----IP-GARCIPMCFEYIYMATYKYLHW 380
QY 299 -----ATALLTLSCGLTLCOSGIYINVLDAIPRYSFIMKASRGFSSIAVYIPYVSGF 353
DB 381 MPICATISLLIGLNLNLSMT--IQVYIVDYTYTYSASALAAATLRSLSFGAFLP-LAGP 436
QY 354 LLSQDPERGMRNVFLLFAVNLGLFYLI-FGEA 387

[illegible]

```

QY 134 -----PEFEYI-----RGVGCVCCLMFVIYDDPSYWMISTS 168
Db 179 MFLALIVFTLPEKLLPKENNTNKAHLDVGILVATTAFTWMLP--ITFNWL--- 234
QY 169 EKEYYITSPL-----KOOGSSKOPLPICAMRSIPMISICGCFSSHOMLVSTWVVYIPT 222
Db 234 ---YMIGALIAIVAVLYIKNMQRPLVNKSFPQNNRYASFLEFIVVMALIQGLYFTFPF 290
QY 223 YISSYHVNIIRNDGLLSALPTIVANVIGNVGYLADFLLTKFRLLTYRKATIILGSPPS 282
Db 291 IMEQYIHQLDPTSLVLPGVGIYAVIVAGLSKGIEYLMSKQAITAIIILALS-ILPA 349
QY 283 -----SALISSLPLYSNGITYTAAIALTLSCGLSTLCOSGYIVNLDIAPRYSFPLMGA 335
Db 350 FAVGNHISLFIVMSIFMGSEFALMAPL-LNEAIRT----IDLMMTGVALFFNYLINV 403
QY 336 SRGFSSIAIVIPYSGFLLSDP-----DEGRNRVFELFAVNLGLPYLLFGEND 388
Db 404 A--VSVCIALAAILIDFKALNPFGNDALSHFG--IILIIGLMISYGLVLEVI----- 454
QY 389 VOEMAKERK 397
Db 454 LNRWTQSEK 462

RESULT 9
W56446
ID W56446 standard; Protein; 2969 AA.
AC W56446:
DT 31-JUL-1998 (first entry)
DE Fragment HGUJ2141 of a new Hepatitis virus g protein.
KW Hepatitis g virus gene; diagnosis; treatment; Hepatitis g virus disease.
OS Hepatitis g virus.
FH Key
FT Misc_difference 1388..1407 Location/Qualifiers
FT FT /note= "nucleotides encoding these residues not given in the specification"
FT FT Misc_difference 1448..1467
FT FT /note= "nucleotides encoding these residues not given in the specification"
PN J10108685-A.
PD 28-APR-1998.
PF 10-AUG-1997; 227387.
PR 10-AUG-1996; JP-227639.
PA (BMLB-) BML KK.
PI MPI: 98-304974/27.
DR N-PSDB: V23080.
PT New hepatitis G virus gene - useful for diagnosing and treating diseases caused by virus
PS Claim 1; Pages 108-112; 128pp; Japanese.
CC The present sequence represents a fragment of a new Hepatitis g virus protein (see also W56441-50 for other fragments). RNA was synthesised from the serum of nine patients judged positive for Hepatitis g virus and cDNA synthesised from this RNA. The cDNA was used as a template in several PCR reactions to isolate fragments of the new gene. The gene CC may be useful for diagnosing and developing treatments for Hepatitis g virus diseases.
SQ Sequence 2969 AA;

Query Match 5.4%; Score 112.5; DB 1; Length 2969;
Best Local Similarity 25.0%; Pred. No. 0.017;
Matches 97; Conservative 35; Mismatches 133; Indels 123; Gaps 23;

QY 79 LPAKSIILG-GGEFATMERGGPPQPERRLCSIALSGMLGCCFPAALLIGRISET-----L 131
Db 637 LPFVSMILGLANLVLYFERMGPO--RLMFVLVIMKLARAFPLALIMG-ISATPGRTSVL 692
QY 132 GMPFVE-----YIFGVGCVCCLMFVVIVDPVSYPM-----ISTSEKEYITSPL 177
Db 693 GAEEFCDFVFEVDTSVLGNV-VASVYAMAAILLSSMSAGCMRHKAIVITYTWCKGT--QAL 749
```

```

QY 178 KOQV-----GSSKKQPIKAMRLSLPISS-----ICLGCF-SHOWLYSTMV 218
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 750 KORVRSPELGGRRPTKPLTIAMCLAST-IMPDAVMLYVGVLTFLGLDADMLAEELV 808
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 YIPYISSVHVHNIIRDNGLLSALPEIY-----AMVIGW-----GGYIAD---F 259
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 809 SRPS-----LRLRLARVECCWAGEKATITRIYSKMCARGAYLFEDHMGSF 853
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 LITKFKRLIYRKATITIGLSPSSALLVSLPYLNSG-YITATALLTLTCSGLSTLCOSGIY 318
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 854 SRAVERL-----LEMDAALPPLSFTPTDCRITIRDAKITLSCG---QCWGLP 898
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 319 INVLIDIAFVRSFIMGASRGFSSIAPIVLPV-----SGF-----LLSODPEFGW 363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 899 V---VAARGVEVLIGVQDVNHLPCPGVPAPAVYIRCGKGFLGVTAKAALTGRDPLHP 954
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 364 RNVEFLLEPAVN-----LLGLLEFLIIG 385
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 955 GNVAVLGTATSRSMGTCINGLLEFTTFHG 982
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10
ID W56450
AC W56450 standard; Protein; 2965 AA.
AC W56450;
DT 31-JUL-1998 (first entry)
DE Fragment F1029 of a new Hepatitis virus g protein.
KW Hepatitis g virus gene; diagnosis; treatment; Hepatitis g virus disease.
OS Hepatitis g virus.
PN J10108685-A.
PD 28-APR-1998.
PF 10-AUG-1997; 227387.
PR 10-AUG-1996; JP-227639.
PA (BMLB-) BML KK.
DR WPI: 98-304974/27.
DR N-PSDB: VZ0384.
PT New hepatitis G virus gene - useful for diagnosing and treating diseases caused by virus
PS Claim 1; Pages 123-127; 128pp; Japanese.
CC The present sequence represents a fragment of a new Hepatitis g virus protein (see also W56441-50 for other fragments). RNA was synthesised from the serum of nine patients judged positive for Hepatitis g virus and cDNA synthesised from this RNA. The cDNA was used as a template in several PCR reactions to isolate fragments of the new gene. The gene may be useful for diagnosing and developing treatments for Hepatitis g virus diseases.
SQ Sequence 2965 AA;

```

Query Match          5.4%; Score 111.5; DB 1; length 2965;
Best Local Similarity 23.2%; Pred. No. 0.02;
Matches 91; Conservative 38; PmiMatches 132; Indels 131; Gaps 22.

QY  LPAKSSILG-GQFAIWERGPPQERSRLCSIALSGMLGCFPAIIIGPISET-----L 131
    ||| ||| :: ||| :: ||| :: ||| :: |||
Db  LRFVSMILGLANLVIFRRMGQ---RLMFVLVIMKLKARCAFPALILMG--ISATRGRTSLV 692

QY  132 GMPFV-----YFGVGVCVCCILMVVYIYDDVSPW-----ISTSEKVIISL 177
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  693 GAEFCEDTVFEVDVTSVGNV-VASVLAAMIALISLSSAGCMRKAVITYTWCkv--QAI 749

QY  178 KOQV-----GSSKQPLDIKAMLSLPIWISICLGCFSSHQWLVSTWVYIPIYISSVHV 230
    : : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db  750 RQRYVSPLEGEPKPLTI-----AM-CLA--SYIWDPAVNLVVAVLLE----- 794

QY  231 NLRDNGLLSALPIVAMVIGWGVLADELTLTKKF--RLITYRKATIIIGSLPSSALIYS 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  794 -----GLFPALD---WA-----LEELLVSRPSLRKLARVVECCVMAGKATIRIAYS 837

QY  289 LPLVNSGY-----ITATALLTSCGSTLQ 314
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  838 KMCARAVYLFDDHMGFSRAVKERILEMDALBPLSTRTDCLRIIPAARTLSCG--QGV 894

```

```

OY      315  SGTYINVLADIPRSSFJMGASRSESSIAPIYPTV-----SGF-----LISODP 359
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      895  MGLPV-----VARQDEVILGIFQDVYNNHLPREFVYTAIVIRRCCKGLSVTKAALIGRDP 950
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      360  EFGMRNVFFLLFAVN-----LLGLLEFLYLIFG 385
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      951  DLHPGNVVVLGATSRSGMTGLNGLLFTTFENG 982
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	11
W56445	
ID	W56445 standard; Protein; 2970 AA.
AC	W56445;
DT	31-JUL-1998 (first entry)
DE	Fragment HGJ1916 of a new Hepatitis virus g protein.
KW	Hepatitis g virus gene; diagnosis; treatment; Hepatitis g virus disease
OS	Hepatitis g virus.
PH	Key
FT	Location/Qualifiers
FT	Misc_difference 2518
FT	/note= "not specified; encoded by CCG"
FT	Misc_difference 2824
FT	/note= "Phe encoded by CTC"
FT	Misc_difference 2857
FT	/note= "Arg encoded by TGC"
FT	Misc_difference 2865
FT	/note= "nucleotides encoding this residue not given"
PN	J10108685-A.
PD	28-APR-1998.
PF	10-AUG-1997; 227387.
PR	10-AUG-1996; JP-227639.
PA	(BMLB-1) BML KK.
DR	WPI: 98-304974/27.
DR	N-PSDB: V23079
PT	New Hepatitis g virus gene - useful for diagnosing and treating
PT	diseases caused by virus
PS	Claim 1; Pages 104-108; 128pp; Japanese.
CC	The present sequence represents a fragment of a new Hepatitis g virus
CC	protein (see also W56441-50 for other fragments). RNA was synthesised
CC	from the serum of nine patients judged positive for Hepatitis g virus
CC	and cDNA synthesised from this RNA. The cDNA was used as a template in
CC	several PCR reactions to isolate fragments of the new gene. The gene
CC	may be useful for diagnosing and developing treatments for Hepatitis g
CC	virus diseases.
CC	Sequence 2970 AA;

Query Match	5.4%	Score 11.5	DB 1	Length 2970
Best Local Similarity	23.0%	Pred. No. 0.022		
Matches	90	Conservative	38	Mismatches 133; Indels 131; Gaps 21;
Qy 79	LPAKSSILG-GQFAIWERMGPOERSRLCSIALSGMLGCFAILGGRISPT-----L	131		
	: : : : : : :			
Db 637	LEPVSSITLGLANLVYLFKRMGPQ---RLMFVIMKILAKAFALPALLMG-ISAITRGRTSVL	692		
Qy 132	GMPEVF-----YIFGGVGCCLLMFVITYDDPVSYPW-----ISTSEKEYIISL	177		
	: : : : : : :			
Db 693	GAPEFCDFVFEYDTSVLGNV-VASYVAMAILLSSMSAGMFKHAKIVYRTCKGK--QAL	749		
Qy 178	KQOV-----GSSKQPLFKAMLSLPLWISICLGCSHQMLVSTINWVYIPIYISSVHV	230		
	: : : : : : :			
Db 750	ROKVAVSPLEGSPAKPLTI-----AM-CLA-SYIPDAVWLNVVGLLF---	794		
Qy 231	NIRDNCLLSALPFIYAVWIGVGYLADPLLRKF--RLITYRKATIILGSLPSSALIYS	288		
	: : : : : :			
Db 794	-----GLEFDALD--WA-----LEELLVSRPSLRKLARVECCVMAGEAKATYRLVS	837		
Qy 289	LPLNLSGY-----ITATALLTSLGSLTIQ	314		
	: : : : : : :			
Db 838	KMCARAYLEFDHNGFSRAVKEHLEMDAALPLSFTRTDCLIRDAARTLSCG---QCV	894		
Qy 315	SGIYINVLDIAPRYSSEFLMGASRGSSIAPIVYPT-----VSGFLISQDP	359		
	: : : : : : :			
Db 895	MGIPV-----VARGDEVLVLGLQDVNHLHPGPEVAPAPVAVIRMGCGGSLGVTBRAALIGRP	950		


```

QY 231 NIRDNLGSLALPFIYAVMWGVGLADFLTKRF--RLITVKRIATIGSLPSSALIYS 288
      || || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      -----GLFDALD-----WA-----LEELLVSRPSLRRLARVECCVMAKERATTIRLVS 837
Db 794 -----GLFDALD-----WA-----LEELLVSRPSLRRLARVECCVMAKERATTIRLVS 837
QY 289 LPLVNGY-----ITATALLTSCGLSTLCO 314
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      838 KMCARGAYLEDMGSEFSRAVKERLLEMDALXELSFTRDCRIIRDAARTLSCG---OCV 894
QY 315 SCGIYINVLDIAPRYSFSLMGASRGFSIAPIVPTV-----SGF-----LLSDPP 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      895 MGLPV-----VARGDEVILGVQDVNHHLPFGVPTAPVPIRRCGKGLGVTKAALTGNDP 950
QY 360 EFGMRNVFLLFAVN-----LLGLFLYLIFG 385
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      951 DLHPGNVWVLGTATSRSGTCLNGLLFTTFHG 982
Db 951 DLHPGNVWVLGTATSRSGTCLNGLLFTTFHG 982

RESULT 14
R95020 standard; Protein: 3011 AA.
AC R95020;
DE Hepatitis GB virus (HGBV) contig C protein prod.
KW Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW reagents; non-A; non-B; non-C; non-D; non-E; clone; GB contig C;
KW tamarin; infected plasma; lambda phage; cDNA library.
OS Hepatitis GB virus.
FH Key
FT misc_difference 1.3011
FT /note= "others correspond to degenerate or STOP
FT codons in T04247"
PD WO9521922-A2.
PD 17-AUG-1995.
PF 14-FEB-1995; U02118.
PR 14-FEB-1994; US-196030.
PR 13-MAY-1994; US-242654.
PR 29-JUL-1994; US-283314.
PR 23-NOV-1994; US-344190.
PR 23-NOV-1994; US-344185.
PR 27-JAN-1995; US-344557.
PA (ABBO ) ABBOTT LAB.
PI BuJk SL, Dawson GJ, Deal SM, Erker JC, Leary TP,
PI Muerhoff AS, Mushahwar IK, Pilot-Matias TU, Schlauder GG;
PI Simons JN;
DR WPI: 95-283123/38.
DR N-PSDB; T04247.
PT Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
PT for diagnosis and therapy of hepatitis GB virus
PS Example 18; Pages 458-471. 661pp. English.
CC Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC infected tamarin plasma, using standard procedures, was used to
CC prepare a lambda phage HGBV cDNA library. Clones were rescued
CC from the lambda phage, searched against a sequence database and
CC found to be unique HGBV sequences. The clones were then used to
CC assemble the sequences GB contig A and B, which were amplified to
CC using random primers. The prod. of which was amplified to give a
CC fragment of GB contig C, then using specified primers the
CC complete sequence of GB contig C was assembled to give T04247,
CC which encodes the proteins R95020-21 (the 3 possible coding strand
CC reading frames). Reagents which comprise the HGBV DNA, or its
CC protein prod. can be used for the diagnosis, therapy or in a
CC vaccine to prevent HGBV infection.
SQ Sequence 3011 AA;

```

Query Match 5.2%; Score 108.5; DB 1; Length 3011;
 Best Local Similarity 23.5%; Pred. No. 0.046;
 Matches 92; Conservative 37; Mismatches 132; Indels 131; Gaps 23;

QY 79 LPAKSSILG-GQFAIWERGPPQERSRSLCSIALSGMLGCFATILIGFISET-----L 131
 Db 753 LPEFVSMTILGLANLVYFRMGPG--RLMPLVWKILARGAFPLALDNG-ISATNGRTSVL 808

```

QY 132 GMEFVE-----YIFGVGCVCCLIMFVYIDPVSYPW-----ISTSEKEYISSL 177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      809 GAEEFEDVVEEDTSLGVN-VASVYAMAIALSSMSAGMKKATIIYRTWCKGY--QXL 865
QY 178 KOOV-----GSSKQPLPKAMRLSLPIIMSTICGFSQHWLVSTMVNYIPIYISSVHV 230
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      866 RQRVVASPLGEGRPKPLI-----AM--CLA--STIMPDVWLVVVAWVLF----- 910
QY 231 NIRDNLGSLALPFIYAVMWGVGLADFLTKRF--RLITVKRIATIGSLPSSALIYS 288
      || || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      910 -----GLFDALD-----WA-----LEELLVSRPSLRRLARVECCVMAKERATTIRLVS 953
QY 289 LPLVNGY-----ITATALLTSCGLSTLCO 314
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      954 KMCARGAYLEDMGSEFSRAVKERLLEMDALXELSFTRDCRIIRDAARTLSCG---OCV 1010
QY 315 SCGIYINVLDIAPRYSFSLMGASR-----GFSIAPIVPTV--SGF-----LLSDPP 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      1011 MGLPV-----VARGDEVILGVQDVNHHLPFGVPTAPVPIRRCGKGLGVTKAALTGNDP 1066
QY 360 EFGMRNVFLLFAVN-----LLGLFLYLIFG 385
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      1067 DLHPGNVWVLGTATSRSGTCLNGLLFTTFHG 1098
Db 1067 DLHPGNVWVLGTATSRSGTCLNGLLFTTFHG 1098

RESULT 15
R99353
R99353 standard; Protein: 470 AA.
AC R99353;
DE 16-DEC-1996 (first entry)
KW Human amine transporter.
KW Amine transporter; neurotransmitter; neurological disease;
KW Alzheimer disease; Parkinson disease; therapy; diagnosis;
KW agonist; antagonist.
OS Homo sapiens.
FH Key
FT misc_difference 402
FT /note= "unidentified amino acid"
PD WO9627009-A1.
PD 06-SEP-1996.
PF 01-MAR-1995; U02645.
PR 01-MAR-1995; WO-U02645.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Cao L, Li Y, Rosen CA;
DR WPI: 96-41275/41.
DR N-PSDB; T35417.
PT Human amine transporter (HAT), stimulates uptake of
PT neuro:transmitter(s) - useful for treating and diagnosing
PT neurological disease, e.g. Parkinson's or Alzheimer's disease.
PS Claim 1; Page 39-41; 58pp. English.
CC Human amine transporter (R99353) stimulates uptake of
CC neurotransmitters. Its amino acid sequence was deduced from
CC a cDNA clone (T35417) isolated from cDNA library derived from
CC a human adrenal gland tumour. Recombinant amine transporter can
CC be produced in prokaryotic or eukaryotic host cells, e.g. E. coli.
CC COS and Spodoptera frugiperda Sf9 insect cells. It may also be
CC used to screen for (ant)agonists of amine transporter activity.
CC Agonists can be used to treat disorders, such as Parkinson's and
CC Alzheimer's diseases, related to under-expression of amine
CC transporter activity. Antagonists (including soluble amine
CC transporter and antibodies) can be used to treat conditions related
CC to excess amine transporter, such as epilepsy, depression,
CC schizophrenia, cognitive impairment, anxiety and migraine.
SQ Sequence 470 AA;

```

Query Match 5.2%; Score 107.5; DB 1; Length 470;
 Best Local Similarity 24.4%; Pred. No. 0.0039;
 Matches 64; Conservative 37; Mismatches 100; Indels 61; Gaps 11;

QY 101 ERSRLCSIALSGMLGCFATILIGFISETLG--WPEVYIF-----GVCVCCLIMFV 154
 Db 224 ERGRAMGTALGIALGLVAGAFSGSVYKEFVSGSAPLLIATLALDGAALQDILGPSKV 283

OY 155 IYDDPVSYPMISTSEKEYIISLKOQVSS-KOPLPIKAMLRSLPIWSICIGCFSHOWLY 213
Db 284 SPESAKGTPLFMLKDPYILVA-----AGSSCFANMGVAILEPTLPIMQTMG-SPKWOL 338
OY 214 STWVYIIPYIYISSVYHVNIIRDNGILSALPPIVAMVIGM-VGYIADFLLTFRKFRLLTVRK 272
Db 339 G-----LAFLPASVSYLIGTNLFGVLANKM-----GRW 366
OY 273 IATILGSL-PSSALIVSLPYLNSGYITATALLTSLGSLTLCOSGIYINVLDIAPRYSS 330
Db 367 LCSLIGMLVVGTSILCVLAHKNFGLIGPNAGLGLXIGWESSMMPIMSHLVD--PRHTS 424
OY 331 -----FLMGASRGFS 340
Db 425 VYGSVHAIDVAFCMGFATIGYS 446

Search completed: May 25, 2000, 14:11:15
Job time: 2352 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 13:38:17 ; Search time 40.12 Seconds

(without alignments)
144.302 Million cell updates/sec

Title: US-09-391-958-1

Sequence: 1 MOWDETLPKRVPSLCSARY.....LIFGEADVQEMAKRKRLRL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: /cgn2_6/pdata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/pdata/1/1aa/6.COMB.pep:*
4: /cgn2_6/pdata/1/1aa/FCRUS.COMB.pep:*
5: /cgn2_6/pdata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2076	100.0	401	2	US-08-805-118-1
2	1991.5	95.9	480	2	US-08-724-394A-11
3	964	46.4	467	2	US-08-805-118-3
4	957.5	46.1	480	2	US-08-724-394A-9
5	722.5	34.8	470	2	US-08-724-394A-10
6	479	23.1	560	1	US-08-647-484-2
7	479	23.1	560	1	US-08-647-481-2
8	479	23.1	560	1	US-08-430-033A-2
9	479	23.1	560	4	PCT-US96-05792-2
10	475	22.9	560	2	US-08-805-118-4
11	122.5	5.9	462	2	US-08-898-976-2
12	122.5	5.9	462	2	US-08-898-976-4
13	106.5	5.1	2873	1	US-08-466-033-15
14	106.5	5.1	2873	1	US-08-638-911A-2
15	106.5	5.1	2873	2	US-08-444-733-15
16	106.5	5.1	2873	2	US-08-464-134-15
17	106.5	5.1	2873	2	US-08-461-361-15
18	106.5	5.1	2873	2	US-08-485-910-15
19	106.5	5.1	2873	4	PCT-US95-06266-15
20	105	5.1	426	5	5268463-8
21	104.5	5.0	470	1	US-08-471-496-2
22	104.5	5.0	470	2	US-08-894-840-2
23	104.5	5.0	2910	1	US-08-466-033-183
24	104.5	5.0	2910	2	US-08-444-733-183
25	104.5	5.0	2910	2	US-08-464-134-183
26	104.5	5.0	2910	2	US-08-461-361-183
27	104.5	5.0	2910	2	US-08-485-910-183
28	104.5	5.0	2910	4	PCT-US95-06266-157
29	103.5	5.0	488	2	US-08-928-692-10

30	103.5	5.0	514	1	US-08-063-552-13	Sequence 13, Appl
31	103.5	5.0	514	4	PCT-US93-05704-13	Sequence 13, Appl
32	101.5	4.9	537	2	US-08-647-397-2	Sequence 2, Appl
33	101	4.9	456	5	5432081-7	Patent No. 5432081
34	98	4.7	457	2	US-08-882-704A-6	Sequence 6, Appl
35	97.5	4.7	457	5	5268463-7	Patent No. 5268463
36	96	4.6	515	1	US-08-063-552-4	Sequence 4, Appl
37	96	4.6	515	4	PCT-US93-05704-4	Sequence 4, Appl
38	94.5	4.6	521	1	US-08-063-552-2	Sequence 2, Appl
39	94.5	4.6	521	4	PCT-US93-05704-2	Sequence 2, Appl
40	94	4.5	3287	2	US-08-477-451-7	Sequence 7, Appl
41	93	4.5	595	2	US-08-677-049-11	Sequence 11, Appl
42	92.5	4.5	465	1	US-08-471-496-9	Sequence 9, Appl
43	92.5	4.5	465	2	US-08-894-840-9	Sequence 9, Appl
44	92.5	4.5	580	2	US-08-677-049-12	Sequence 12, Appl
45	92	4.4	397	1	US-08-098-141-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-805-118-1
; Sequence 1, Application US/08805118
; Patent No. 5985604
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
; TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,118
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0221 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRA1UT02
; CLONE: 754412
; US-08-805-118-1

Query Match 100.0%; Score 2076; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 9.5e-212;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOVDELTPRKVPKSLCSARYGIALVLFHCNFTTIAQNVIMNTVMAMVNSTSPQSOLNDS 60
DB 1 MOVDELTPRKVPKSLCSARYGIALVLFHCNFTTIAQNVIMNTVMAMVNSTSPQSOLNDS 60
QY 61 SEVLVPVDSRGSGSKAPKSLPAKSSILGGCFALWERNKGPQERSRLCSIALSGMLGCFPA 120
DB 61 SEVLVPVDSRGSGSKAPKSLPAKSSILGGCFALWERNKGPQERSRLCSIALSGMLGCFPA 120
QY 121 ILIGGFISSETLGMPEFVFYIFGSGVCVCCILMFVYIYDDPVSPWISTSEKEYIISLKQ 180
DB 121 ILIGGFISSETLGMPEFVFYIFGSGVCVCCILMFVYIYDDPVSPWISTSEKEYIISLKQ 180
QY 181 VSSSQOPLPIKMLNSLPWISICLCFSGHQLVSTMVVIIPYIISVYHVNIRONGLLSA 240
DB 181 VSSSQOPLPIKMLNSLPWISICLCFSGHQLVSTMVVIIPYIISVYHVNIRONGLLSA 240
QY 241 LPFIYAWYIGMWGYLADFLTKRRLTVRKIATILGSLPSALIVSLPYINSGYTAT 300
DB 241 LPFIYAWYIGMWGYLADFLTKRRLTVRKIATILGSLPSALIVSLPYINSGYTAT 300
QY 301 ALLTSLGSLTLCQSGIYINVDIAPRYSSFLMGASRGSSIAPIYIVPVSGFLLSODPE 360
DB 301 ALLTSLGSLTLCQSGIYINVDIAPRYSSFLMGASRGSSIAPIYIVPVSGFLLSODPE 360
QY 361 FGRNVFFLLFAVNLLGLFYILFGEADVQEWAKERKRLRL 401
DB 361 FGRNVFFLLFAVNLLGLFYILFGEADVQEWAKERKRLRL 401

RESULT 2
US-08-724-394A-11
Sequence 11, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kromal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREM LLP
STREET: Two Embarcadero Center, 8th floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Flits, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant

MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Region
LOCATION: 1..480
OTHER INFORMATION: /note= "NP74"
US-08-724-394A-11

Query Match 95.9%; Score 1991.5; DB 2; Length 480;
Best Local Similarity 82.9%; Pred. No. 1.1e-202;
Matches 398; Conservative 1; Mismatches 2; Indels 79; Gaps 3;

QY 1 MOVDELTPRKVPKSLCSARYGIALVLFHCNFTTIAQNVIMNTVMAMVNSTSPQSOLNDS 60
DB 1 MOVDELTPRKVPKSLCSARYGIALVLFHCNFTTIAQNVIMNTVMAMVNSTSPQSOLNDS 60
QY 61 SE-----VLPVDSFGLSKA 75
DB 61 SEXXXVLPVDSFGLSKA 120
QY 76 PKSLP-----AKSSILGGCFALWERNKGPQER 102
DB 76 PKSLPXXAKSSILGGCFALWERNKGPQER 180
QY 103 SRLCSIALSGMLGCFETAILIGGFISSETLGMPEFVFYIFGSGVCVCCILMFVYIYDDPVSY 162
DB 103 SRLCSIALSGMLGCFETAILIGGFISSETLGMPEFVFYIFGSGVCVCCILMFVYIYDDPVSY 240
QY 163 PWISTSEKEYIISLKQOVSSKOPPIKAMRLSLPWSICLCFSGHQLVSTMVVIIPYI 222
DB 163 PWISTSEKEYIISLKQOVSSKOPPIKAMRLSLPWSICLCFSGHQLVSTMVVIIPYI 300
QY 223 YISSYHVNIRONGLLSALPFIYAWYIGMWGYLADFLTKRRLTVRKIATILGSLP 281
DB 223 YISSYHVNIRONGLLSALPFIYAWYIGMWGYLADFLTKRRLTVRKIATILGSLP 360
QY 282 SSALIVSLPYINSGYITPTATLITSCGLSTLCQSGIYINVDIAPRYSSFLMGASRGSS 341
DB 282 SSALIVSLPYINSGYITPTATLITSCGLSTLCQSGIYINVDIAPRYSSFLMGASRGSS 420
QY 342 IAPVIVPVSGFLLSODPEFGMRNVFFLLFAVNLLGLFYILFGEADVQEWAKERKRLRL 401
DB 342 IAPVIVPVSGFLLSODPEFGMRNVFFLLFAVNLLGLFYILFGEADVQEWAKERKRLRL 480

RESULT 3
US-08-805-118-3
Sequence 3, Application US/08805118
Patent No. 5985604
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,118
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0221 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 467 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 450532
; US-08-805-118-3

Query Match      46.4%; Score 964; DB 2; Length 467;
Best Local Similarity 41.3%; Pred. No. 7.4e-94;
Matches 193; Conservative 68; Mismatches 140; Indels 66; Gaps 3;

QY 1 MOVDELIRKRVPSLCSARYGIALVLFPCNFTIAQNVINIMVAMVNSTSPQSOLNDS 60
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MQNDNRLPKRVKVFCSFRIGLSFLVHCNVITTAQACINLTVMVAVNSTDPHGLPNTS 60
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 SEVL-----PVDSFGGLSKAPKSL-----80
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TKRLDINIKPMYMWSPDIOGILSSTSYGVIIIQVPVGFSGIYSTKMKMIGALCLSSV 120
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 80 -----PAKSSILGQFAIMERWGPPOERSRLCSIALSGML 115
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 LSLIPPAAGIGVAVVYVCAVAGAAQIVATQAQFEIYVMAAPLERGLTSMSTGFL 180
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 116 GCFTALIGFSETLGMPVEYIFGGVGCVCCLLMFVVYIDDPVSPATSTSEKEIIS 175
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 GPTVLVLTGVICESLQMPVFIIFGACGAVCLLMFVLEDDPKDKPCISISEKEIIS 240
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 176 SLKQGVSSKQPLPIKAMLSLPIWSICLGCFSHOMLVSTMVVYIPTYISSVYHVNIRN 235
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 SLVQGVSSSQSLPIKALIKSLPVMAISIGSFTEFWSHNMILTYPFIMSLMHNKEN 300
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 236 GLLSALPIFYAVYIGVAGYLADFLTKK-FRLITVRKIAITIGSLSSALIVSLPYLNS 294
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 GFLSSLPYLFAMWICGNLAGQSDFFLRNLTSLVAVRKLTAAAGFLPAIFGVCLPYLSS 360
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 295 GYTATALLTSCGSLTSCGSIYINVDIAPRYSFPMGASRGFSSIAPIVYPTVSGFL 354
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 TFSYIYIFLLAGATGSCFCLGVFINGLDIAPRYFGFIKACSTLTGMLIAGLSTLTGLI 420
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 355 LSQDPFGWRNVFFLLFAVNLGLFLYLFGEADVQEMAKERKLTRL 401
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 421 LKQDPESAMFKFTILMAAINVTGLIFLYIVATFIQDMAKEKQHTRL 467
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
; US-08-724-394A-9
; Sequence 9, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kironmal, Gregory S.
; APPLICANT: Laufer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolf, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereo
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..480
; OTHER INFORMATION: /note="NPT1"
; US-08-724-394A-9

Query Match      46.1%; Score 957.5; DB 2; Length 480;
Best Local Similarity 40.2%; Pred. No. 3.8e-93;
Matches 193; Conservative 68; Mismatches 140; Indels 79; Gaps 3;
```

```

QY 1 MOVDELIRKRVPSLCSARYGIALVLFPCNFTIAQNVINIMVAMVNSTSPQSOLNDS 60
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MQNDNRLPKRVKVFCSFRIGLSFLVHCNVITTAQACINLTVMVAVNSTDPHGLPNTS 60
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 SEVL-----PVDSFGGLSKA 75
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TKRLDINIKRXXXXXXXXXXXXPMYMWSPDIOGILSSTSYGVIIIQVPVGFSGIYST 120
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 76 PKSL-----PAKSSILGQFAIMERWGPPOER 102
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 KKMIGFALCLSSVLSLIPPAAGIGVAVVYVCAVAGAAQIVATQAQFEIYVMAAPLER 180
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 103 SRLCSIALSGMLGCFPAITIGFISSETLGMPVEYIFGGVGCVCCLLMFVVYIDDPVSY 162
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 GRUFSMTSGFLGPTFLVLTGVICESLQMPVFIIFGACGAVCLLMFVLEDDPKDH 240
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 163 PWISTSEKEYIISLQGVSSKQPLPIKAMLSLPIWSICLGCFSHOMLVSTMVVYIPT 222
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 PCISISEKEYITSSLVQGVSSRSQSLPIKALIKSLPVMAISIGSFTEFWSHNMILTYPM 300
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 223 YISSVYHVNIRDNGLSALPIFYAVYIGVAGYLADFLTKK-FRLITVRKIAITIGSLP 281
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 FINSMLHVNIRKENGFLSSLPYLFAMWICGNLAGQSDFFLRNLTSLVAVRKLTAAAGFL 360
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 282 SSALIVSLPYLNSGYITATALLTSCGSLTSCGSIYINVDIAPRYSFPMGASRGFSS 341
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 PAIFGVCLPYLSSPYFIVIFLLAGATGSCFCLGVFINGLDIAPRYFGFIKACSTLTGM 420
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 342 IAPYIVPVSGLISQDPFGWRNVFFLLFAVNLGLFLYLFGEADVQEMAKERKLTRL 401
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 421 IGLIAGLSTLTGLIKQDPESAMFKFTILMAAINVTGLIFLYIVATFIQDMAKEKQHTRL 480
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

[illegible]

		23.13	Score 479;	DB 1;	Length 560;
	Query Match Similarity	28.22;	Prod. No. 2.3e-42;		
	Best Local Similarity	28.22;	Prod. No. 2.3e-42;		
	Matches 112; Conservative	66;	Mismatches 135;	Indels 62;	Gaps
QY	12	VPSLCSARYGIALYALHFCNFTTIAQNYINMIVTVAWVNSTSPQSLNDSEVLPVDSFG	71		
Db	161	IPSAARVHVGCVI-----FVRIQLGVEEVTPAC-----	191		
QY	72	LSKARKSLPAKSSILGQFAIEMRWGPQDERSICALSGLMLCCFTAILIGFISFTL	131		
Db	191	-----HGIMSKMAEPLEERSKLATTAACGSGYAAVAVAMPPLAGIVQYS	232		
QY	132	GMPEFYITFGGVCYCCCLIMPFVYIYDDPVSYPIKSTSEKEYIISLTKQOVSSKO---PL	188		
Db	233	GMSVFFIYVYSGFGFPIWTFMLWLSYESPRLAHPISISEEKRYI---EALIESAKLNKPL	288		

```

Oy 189 ----PIKAMLSCLPMSICLGCFSHOWLVTSMVYIPYISVYVNIIRDNGLSALPF 243
Db 269 TRFSFPMRRFFSMMVYAIIVANFCRSTFYLLLSQDPYFEEVEFEIISKVLVSALPH 348
Oy 244 IVAMVIGWGGYLADELLTKFRLLIT-VKRIATILSLSPSALIYSLPYLNSGYIATATL 302
Db 349 LVMITIVIPGIGIADFLBSRIMSTTNVRKLNCCGFCGEATLLLVGISHSKGV-AISF 407
Oy 303 LTLSCGLSTLCOSGIYINVLADIAPRYSFFLMGASRGESSIADPYVPTVSGFLLSQDPFG 362
Db 408 LVLAAGFGSFAISGPNVNHLDIAPRYASIIAMGINSVGFLSGWCPYIIGAMTKHRTRE 467
Oy 363 WRNVFELFAVNLGLFLFLGCEADVQEMAKERKIT 399
Db 468 WOYFELIASLVHYGVIFGVFASGEKOPMAPEEMMS 504

RESULT 7
US-08-647-481-2
? Sequence 2, Application US/08647481
? Patent No. 5618918
? GENERAL INFORMATION:
? APPLICANT: NI, Binhui
? APPLICANT: Paul, Steven M.
? TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
? TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
? NUMBER OF SEQUENCES: 3
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Eli Lilly and Company
? STREET: Lilly Corporate Center
? CITY: Indianapolis
? STATE: Indiana
? COUNTRY: United States of America
? ZIP: 46285
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/647,481
? FILING DATE: 14-MAY-1996
? CLASSIFICATION: 536
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/430,033
? FILING DATE: 27-APR-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Gaylo, Paul J.
? REGISTRATION NUMBER: 36,808
? REFERENCE/DOCKET NUMBER: X-10006
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (317) 276-0756
? TELEFAX: (317) 276-3861
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 560 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-647-481-2

Query Match 23.1%; Score 479; DB 1; Length 560;
Best Local Similarity 28.2%; Pred. No. 2.3e-42;
Matches 112; Conservative 68; Mismatches 155; Indels 62; Gaps 7;

Oy 12 VPSLCSARIGIALVLFHCFTTIAQNVINNTIMVAVNVNSTSPQSQNDSEVLVDSFG 71
Db 161 IPSARVHYGCIY-----FVRILGLVGEVGYTPAC----- 191
Oy 72 LSKARKSLPAKSSILGGQFAIWERGPPQERSRLCSIALSGMLGCFALILGCFISETL 131
Db 191 -----HGIMSKWAPPLERSRLATTAFCGSIAGAVVAVMPLAGVLYQVS 232

```

Db 191 -----HGIMSKWAPPLERSRLATTAFCGSGYAGAVVAMPPLAGLVQYS 232

QY 132 GMPFVYIFGVCVCCLLMPVYIDDPVSYPMWISTSEKEYIISLKQOVSSKO---PL 188

Db 233 GWSVFYVYGGSGIFWYLFMWLVSYESPALHPSISEEKRTY---EPAIGSAXLMPPL 288

QY 189 -----PIKAMRLSPFWSICLGCFSHOWLVSTMVYIPTYISSYHVNIIRDNGLSALPF 243

Db 289 TKFSPTMRFPTSMVYAIIVANFCRSMWTFYLLISQPDYEEVGFELSKVGLVSALPH 348

QY 244 IVAWYIGVGYLADFLTKKFRILT-VKTIATILGSLPSSALIVSLPLNSGYTTATL 302

Db 349 LVMTIIVIGGQIADFLRSRIMSTYNVKIKMNCGFGMEATLLLVGYSHSKGV-AISF 407

QY 303 LTLSCGLTLCOSGIYINVDIAPRYSFIMGASRGFSSIAPIVYPTVSGFLISODPEFG 362

Db 408 LVLAVFGSGFALSGFNVHLDIAPRYASILMGISNGVETLSGMWCPPIVGAMTKHKTREE 467

QY 363 WRNVFLLFAVNLGLLFYLIFFGADVOEMAKERKLT 399

Db 468 WQYVFLIASLVHYGVIIFYGVFASGEKQPMAPPEMS 504

RESULT 9

PCT-US96-05792-2

: Sequence 2, Application PC/TUS9605792

: GENERAL INFORMATION:

: APPLICANT: NI, Binhui

: APPLICANT: Paul, Steven M.

: TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC

: TITLE OF INVENTION: PHOSPHATE COTRANSPROTHER

: NUMBER OF SEQUENCES: 3

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Eli Lilly and Company

: STREET: Lilly Corporate Center

: CITY: Indianapolis

: STATE: Indiana

: COUNTRY: United States of America

: ZIP: 46285

: COMPUTER READABLE FORM:

: MEDIUM TYPE: floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patentin Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: PCT/US96/05792

: FILING DATE:

: CLASSIFICATION:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 08/430,033

: FILING DATE: April 27, 1995

: ATTORNEY/AGENT INFORMATION:

: NAME: Gaylo, Paul J.

: REGISTRATION NUMBER: 36,808

: REFERENCE/DOCKET NUMBER: X-10006

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (317) 276-0756

: TELEFAX: (317) 276-3861

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 560 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

PCT-US96-05792-2

Query Match 23.1%; Score 479; DB 4; Length 560;

Best Local Similarity 28.2%; Pred. No. 2.3e-42;

Matches 112; Conservative 68; Mismatches 155; Indels 62; Gaps 7;

QY 12 VPSLCARYGIALVLRHCNFTTIAQNVIMNITVAMVNSTSPQSQINDSSEVLPAVDSFG 71

Db 161 IPSARVHRCYI-----FVRILOGLVGYTPAC-----191

QY 72 LSKAPKSLPANSILGOFALWERNWPPQERSRLCSIALSGMLACCTAIIIGFISRTL 131

Db 191 -----HGIMSKWAPPLERSRLATTAFCGSGYAGAVVAMPPLAGLVQYS 232

QY 132 GMPFVYIFGVCVCCLLMPVYIDDPVSYPMWISTSEKEYIISLKQOVSSKO---PL 188

Db 233 GWSVFYVYGGSGIFWYLFMWLVSYESPALHPSISEEKRTY---EPAIGSAXLMPPL 288

QY 189 -----PIKAMRLSPFWSICLGCFSHOWLVSTMVYIPTYISSYHVNIIRDNGLSALPF 243

Db 289 TKFSPTMRFPTSMVYAIIVANFCRSMWTFYLLISQPDYEEVGFELSKVGLVSALPH 348

QY 244 IVAWYIGVGYLADFLTKKFRILT-VKTIATILGSLPSSALIVSLPLNSGYTTATL 302

Db 349 LVMTIIVIGGQIADFLRSRIMSTYNVKIKMNCGFGMEATLLLVGYSHSKGV-AISF 407

QY 303 LTLSCGLTLCOSGIYINVDIAPRYSFIMGASRGFSSIAPIVYPTVSGFLISODPEFG 362

Db 408 LVLAVFGSGFALSGFNVHLDIAPRYASILMGISNGVETLSGMWCPPIVGAMTKHKTREE 467

QY 363 WRNVFLLFAVNLGLLFYLIFFGADVOEMAKERKLT 399

Db 468 WQYVFLIASLVHYGVIIFYGVFASGEKQPMAPPEMS 504

RESULT 10

US-08-805-118-4

: Sequence 4, Application US/08805118

: Patent No. 5985604

: GENERAL INFORMATION:

: APPLICANT: Lal, Preeti

: APPLICANT: Bandman, Olga

: TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT

: TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER

: NUMBER OF SEQUENCES: 7

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Incyte Pharmaceuticals, Inc.

: STREET: 3174 Porter Drive

: CITY: Palo Alto

: STATE: CA

: COUNTRY: US

: ZIP: 94304

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM compatible

: OPERATING SYSTEM: DOS

: SOFTWARE: FastSeq Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/805,118

: FILING DATE: Filed Herewith

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER:

: FILING DATE:

: ATTORNEY/AGENT INFORMATION:

: NAME: Billings, Lucy J.

: REGISTRATION NUMBER: 36,749

: REFERENCE/DOCKET NUMBER: PF-0221 US

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 415-855-0555

: TELEFAX: 415-845-4166

: INFORMATION FOR SEQ ID NO: 4:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 560 amino acids

: TYPE: amino acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: IMMEDIATE SOURCE:

: LIBRARY: GenBank

CLONE: 507415

US-08-805-118-4

Query Match 22.9%; Score 475; DB 2; Length 560;
Best Local Similarity 28.0%; Pred. No. 6,1e-42;
Matches 11; Conservative 66; Mismatches 158; Indels 62; Gaps 6;

QY 12 VPSLCSARGLALVLFHFCFTTIAOVNIMTMVAMVNSTSPQSLDNDSSEVLVPVDSFG 71
DB 161 IFSARVHAGCVI-----FVRILOGLVGVTYPAC-----191
QY 72 LSKARKSLPAKSSILGOGFAIWERMGPPQERSRLCSIALSGMLGCTAILIGGFISETL 131
DB 191 -----HGIMSKNAPLERSLRTATACGSAVAVMPLAGLVQYS 232
QY 132 GMPFPIYFGVGCVCCLMFVYIYDDPVSPWISSEKEEYIISLKQOVSS-----185
DB 233 GMSVFEYVYGSFGIFWYLFMLVSESPALHPSISEEERKYI---EDAHGESAKLMNPV 288
QY 185 -KQPLPIKAMLSLPIWISICGCFSHQWLVTMYYVYIPYIISVYHVNRDNGLSALPF 243
DB 289 TKFNPWKRFPFSPMPYALITVANFCRSMTFYLLISQPAFEEVEGFEESKVGLSALPH 348
QY 244 IYAWYIGWVGYLADFLTKKRLIT-VKRIATILGSLPSSALIYSLPYLNSGYITATL 302
DB 349 LVMITIVPIGGIADFLBSRHIMSTTNVRKLMNCGFGEMATLLLVG:SHSKGV-AISF 407
QY 303 LTLSCGLSTLCSGIYINVLDIAPRYSFLMGASRGFSSIAPIVPTVSGFLLSODPERG 362
DB 408 LVLAVGFSGFAISGFVNHLDIAPRYASILMGISNGVGLSGVCPILVIGAMTKHKTRE 467
QY 363 WRNVFELFAVNLGLFLFLIGLEADVQWAKERKIT 399
DB 468 WQYVFLASLVHGYGVIFGVFASGEKQPMAPPEEMS 504

RESULT 11

US-08-898-976-2

Sequence 2, Application US/08898976

Patent No. 5891670

GENERAL INFORMATION:

APPLICANT: Burnham, Martin

APPLICANT: Lonetto, Michael

APPLICANT: Warren, Patrick

TITLE OF INVENTION: NOVEL TETRACYCLINE RESISTANCE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/898,976

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: GM10044

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 462 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-898-976-2

Query Match 5.9%; Score 122.5; DB 2; Length 462;
Best Local Similarity 17.7%; Pred. No. 0.0001;
Matches 76; Conservative 86; Mismatches 158; Indels 109; Gaps 18;

QY 20 YGIALVLFHFCFTTIAOVNIMN---ITVAMVNSTSPQSLDNDSSEVLVPVDSFGSKAP 76
DB 92 YGIVLIT---FGSIIGNIFQHSPLTLVGRIQTA-----GLAAB 129
QY 77 KSLPAKSSILGOGFAIW-ERMGPPOERSRLCSIALSGMLGCTAILIGGFISETLGW--134
DB 130 T-----LYIYVAKYLSKEDQKTYLGISTSSYLSLYIGTSGFIFTYLHMN 178
QY 134 -----PFVEYI-----FGVGCVCCLMFVYIYDDPVSPWIS 168
DB 179 MELIATIVFTLPEFLKLPKENNTKKAHLDFYGLLVATIAITWLF--ITFNNL---234
QY 169 EKEYIISL-----KQOVSSKQPLPIKAMLSLPIWISICGCFSHQWLVTMYYVYIPT 222
DB 234 ---YMGALIAIIVFALYIKNAORPLVNRKSFQNKRYASFLFVFWVAIQIQYTFPF 290
QY 223 YISSVHVHNRDNGLSALPFITAWYIGWVGYLADFLTKKRLITVKRIATILGSLPS 282
DB 291 IMEQIYHLDPTTSLLVPGYIVAVIVGALSGIGEYLSKQIITATILALSL-ILPA 349
QY 283 -----SALIVSLPYLNSGYITATLTLSCGLSTLCSGIYINVLDIAPRYSFLMGA 335
DB 350 FAVGNHISLFIYISMIFPAGSFALMTAPL-LNEAIKT-----IDLNMTGVALIGYNIITVY 403
QY 336 SRGFSIAPIVPTVSGFLLSODP-----EGWNRVFLFAVNLGLFLFLIGLEAD 388
DB 404 A--VSVGIALAALIDFALNPGDALSHRG--IILIIILGMSIVGLVLEVI-----454
QY 389 VOEWAKER 397
DB 454 LNRWTOSEK 462

RESULT 12

US-08-898-976-4

Sequence 4, Application US/08898976

Patent No. 5891670

GENERAL INFORMATION:

APPLICANT: Burnham, Martin

APPLICANT: Lonetto, Michael

APPLICANT: Warren, Patrick

TITLE OF INVENTION: NOVEL TETRACYCLINE RESISTANCE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/898,976

FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10044
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-898-976-4

Query Match 5.9%; Score 122.5; DB 2; Length 462;
Best Local Similarity 17.7%; Pred. No. 0.0001;
Matches 76; Conservative 86; Mismatches 158; Indels 109; Gaps 18;

QY 20 YGAIALHFCNFTTAQNVIM--ITWYAVNSTPOSQNDSSBVLVDSFGGSKAP 76
DB 92 YGVILIT--FGSTIGNFQHSPLTLVGRITQTA-----GLAANE 129
QY 77 KSLPKKSSILGGQFAIW-ERMGPPOERSRLCSIALSGMLGCFALILGGFISFTLW-- 134
DB 130 T-----LVYIYAKYLSKEDOKTYLGLSTSSYSLSLVIGTLGGCFYFLVHMTN 178
QY 134 -----PFEVYI-----FGVGVCCOLLMFVYVDDPVSYWISIS 168
DB 179 MFLALIVFTLPFLFKLLPKRNNTKKAHLDFVGLILVATTAFTVWL-ITNFWNL--- 234
QY 169 EKEYIISL-----KQOVGSSKQPLPKAMLRSLPWSICIGCFSHQWLVSTWVYIPT 222
DB 234 ---YMGALIAIIVFALYIKNAQRLVKNKSPQNKRYASFLEIVYMAIQIGYFTTFP 290
QY 223 YISSYHNINRNGLLSLPFIYAVNIGWGYLADFLITKFKRLLITVYKTIATLIGSLPS 282
DB 291 IMEQLYHLQDTSLLPVGIVAVIAGLSKIGEYLNKQAITAILIALSL-ILPA 349
QY 283 -----SALIVSLPYLNSGYITATALLTSCGSTLQSGYINVDIARRYSFLMGA 335
DB 350 FAVGNHISLFIYSMTFFAGSFLAMAPL-LNEAIRT-----IDLMTGVALGFYLLITNV 403
QY 336 SRGSSIAPIVPIVYSGFLSQDP-----EFGWRNVFLLFAVNLGLLFYLIFFGEAD 388
DB 404 A--VSQIAIAALIDFKALNFPQNDALSHFG--ILITLIGLSMISGVLLVFI----- 454
QY 389 VOENAKERK 397
DB 454 LNRRTQSEK 462

RESULT 13
US-08-466-033-15
Sequence 15, Application US/08466033
Patent No. 5766840
GENERAL INFORMATION:
APPLICANT: Kim, Jungsuh P.
APPLICANT: Wages, John
APPLICANT: Young, Lavonne M.
APPLICANT: Fiy, Kirk E.
APPLICANT: Linnen, Jeffrey M.
TITLE OF INVENTION: Hepatitis G Virus and Molecular
NUMBER OF SEQUENCES: 277
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,033
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,886
FILING DATE: 15-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/357,509
FILING DATE: 16-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,729
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,271
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,558
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,543
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,985
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2873 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-033-15

Query Match 5.1%; Score 106.5; DB 1; Length 2873;
Best Local Similarity 22.9%; Pred. No. 0.078;
Matches 89; Conservative 42; Mismatches 133; Indels 125; Gaps 22;

QY 79 LPAKSTILG-QGFAIMERNMGPOERSRLCSIALSGMLGCFALILGGFISFT-----L 131
DB 637 LPVYSMTLGLANLVLYRWRMGPO--RLMFLVLMKLRAGAPLALLMG-ISARRGRTSVL 692
QY 132 GWPEVF-----YIRGGVCCOLLMFVYVDDPVSPW-----ISISEKEYIISL 177
DB 693 GAERCFDATFEVDTSVYGWV-VASVYMAIALLSMSAGCKRRKAVIYRTWCQGY-QAT 749
QY 178 KQOVGSSKQPL-----PKAMLRSLPWSICIGCFSHQWLVSTWVYIPTVYSSVYHNIR 233
DB 750 RQRY--VRSPLEGGRPAKPL-----TFWCLA--SYIMPDVAMVVVALVLF----- 794
QY 234 DNGLSALPFIYAVNIGWGYLADFLITKFK--RLITVYKTIATLIGSLPSRLIYSLPY 291
DB 794 --GLFDALD---WA-----LEELIVSRPSLRILARVECCVMAGKATVRLVSKMK 840
QY 292 LNSGY-----ITATALLTSCGSTLQSGI 317

DB 841 ARGATLEDMHGSFRAVAKERLEMDALEPLSFRTDRIIRDAARTLSCG---QCVMGL 897
QY 318 YINVLADIAPRYSFELMGASGFSSIAPIVTPV-----SGF-----LISQDEFG 362
DB 898 PV-----VARRGDEVLLIGFODVNHLPFGFPTAPVPIRRCGKGLGVTKAALGRDPDLH 953
QY 363 WRNVFELFAVN-----LLGLLFYLIFFG 385
DB 954 PGNVNVLGTATSRSMGTCLNGLLFTTFHG 982

RESULT 14
US-08-638-911A-2
; Sequence 2, Application US/08638911A
; Patent No. 5766916
; GENERAL INFORMATION:
; APPLICANT: Belayev, Alexander S.
; APPLICANT: Chong, Susan M.
; TITLE OF INVENTION: Hepatitis G Virus Protease
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/638,911A
; FILING DATE: 04/24/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Evans, Susan T.
; REGISTRATION NUMBER: 38,443
; REFERENCE/DOCKET NUMBER: 4600-0124
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2873 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-638-911A-2

Query Match 5.1%; Score 106.5; DB 1; Length 2873;
Best Local Similarity 22.9%; Pred. No. 0.078;
Matches 89; Conservative 42; Mismatches 133; Indels 125; Gaps 22;

QY 79 LPAKSSILG-GQFAIWERKCPPOERSRLCSIALSGMLGCFYAILIGCFISER-----L 131
DB 637 LPVYASHILGLANLVYLRMLGPQ--RLMFLVLMKLAGAFPLALMG--ISATRGRTSVL 692
QY 132 GMPFVF-----YIFGGVGCVCCLLMFVVIYDDPVSYPM-----ISTSEKYEISSL 177
DB 693 GAFCFCDATEVETSVLGVV--VASVYAMAIALLSMSAGWRKRAVYRTWCKGY--QAI 749
QY 178 KOQVSSKKPL-----PIKAMLRSLPIWSICLGCFSHQMLVSTWVYIPTYISSVYHVNIR 233
DB 750 ROHV--VRSPLGGRPAKPL-----TFAWCLA--SYIWPDAVMMVVVVLVILF----- 794
QY 234 DNGLLSALPPIVAMVIGMGYIADFLITKFF--RLITYRKITITIGSLDSSALIYSLPY 291
DB 794 --GLEFDALD---WA-----LEEILVSRPSLRRLARAVECCVMGEKATYVRLVSKMC 840
QY 292 LNSGY-----ITTATALLTLSCGLSTIGQSGI 317

DB 841 ARGATLEDMHGSFRAVAKERLEMDALEPLSFRTDRIIRDAARTLSCG---QCVMGL 897
QY 318 YINVLADIAPRYSFELMGASGFSSIAPIVTPV-----SGF-----LISQDEFG 362
DB 898 PV-----VARRGDEVLLIGFODVNHLPFGFPTAPVPIRRCGKGLGVTKAALGRDPDLH 953
QY 363 WRNVFELFAVN-----LLGLLFYLIFFG 385
DB 954 PGNVNVLGTATSRSMGTCLNGLLFTTFHG 982

RESULT 15
US-08-444-733-15
; Sequence 15, Application US/08444733
; Patent No. 5824507
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungshuh P.
; APPLICANT: Wages, John
; APPLICANT: Young, Layonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linnen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,733
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2873 amino acids
; TYPE: amino acid

```

;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-444-733-15

```

Query Match	5.1%	Score 106.5	DB 2	Length 2873
Best Local Similarity	22.9%	Pred. No. 0.078		
Matches 89	Conservative 42	Mismatches 133	Indels 125	Gaps 22

```

QY 79 LPKASIIIG-COFALTEWKGQPOERSRCSIALSGMLLGGCFALLIGGFISER-----L 131
Db 637 LPVYSMTIIGLANLVITYFWMIGPQ---RMLFVLTKKLARGAPLALLMG-ISAIRGRTSYL 692
QY 132 GMPFVE-----YIFGGVGCVCCLMFVVIYIDDPYSPW-----ISTSEKIYIISL 177
Db 693 GAEECFDATFEVDITSVLGW-VASVAMALIALSSMGAGGRHKAVIYRWCGY--QAI 749
QY 178 KQGVSSKQPL-----PIKAMLRSLPIISICIGCFSHQWLVSTWVVIYPIYISSVYHNIR 233
Db 750 RQRV-VNSPLGEGRPAPKPL-----TEAWCLIA-SYTMPADVMMVVVALVLP----- 794
QY 234 DNGLLSALPFIYVAVWIGWGVSYLDELITTKF--RLITVYKAITILGSLPSSALLVSLPY 291
Db 794 ---GLEFDALD---WA-----LEEIIYSRSLRLIARVVECCVMAGEKATYRLVLSKMC 840
QY 292 LNSGY-----T-----TEATALLTISGSLTSGSGI 317
Db 841 ARGAVYLPDHMSFSRVAVERKERLEMDALBPLSTRTDCKRIIRDAANTLSG---QCVMEI 897
QY 318 YIYNVLDIAPRYSFSLMGASRGFSSIAFIVPYV-----SGF-----LISQDPEFG 362
Db 898 PV-----VARRDEVLIVGFODVNHLPFGVPFAPVARIROGKGFELVTAKAULGRPDILH 953
QY 363 WRNVEFLLFVAVN-----LLGLLFYLLFG 385
Db 954 PGNAVYLGITATSRSMGTCLNGLLLETFYFHG 982

```

Search completed: May 25, 2000, 14:25:01
Job time: 2804 sec

OM of: US-09-391-958-1 to: GenEmbl.* out_format : pfs
Date: May 25, 2000 1:47 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+ -p2n.model -DEV-x1p
-O/cgcn2_1/USPTO.spool/US09391958/runat.24052000_213330_7161/app-query.fasta.1
-DB-GenEmbl -OPM-fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=Pct -ALIGN=15 -MODE=LOCAL
-OUTFM=pfs -NORM-ext -MINLEN=0 -MAXLEN=1000000 -USER=US09391958
-NCPU=6 -ICPU=3 -NO_XLIPXY -WAIT -THREADS=1

Search information block:

Query: US-09-391-958-1
Query Length: 401
Database: GenEmbl.*
Database sequences: 882769
Database length: -486395729
Search time (sec): 737.170000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_pat:AR036571	+ 2061.00	3007.97	1.7e-159	1780	AR036571 Sequence 19 from pate
gb_p3:HSU90545	+ 2061.00	3007.90	1.7e-159	1795	U90545 Human sodium phosphate
gb_p1:AB020527	+ 1023.50	1483.30	1.4e-74	2626	AB020527 Homo sapiens mRNA for
gb_p2:HSNAP11	+ 964.00	1400.99	5.6e-70	1549	X71355 H.sapiens mRNA for sodi
gb_ro:RNU28504	+ 959.00	1392.80	1.7e-69	1700	U728504 Rattus norvegicus Na+/H
gb_ro:MMP1TCT	+ 951.00	1380.10	8.1e-69	1885	X77241 M.musculus Npt1 mRNA fo
gb_p1:HMARP14	+ 944.00	1370.30	2.8e-68	1794	D28512 Human mRNA for renal Na
gb_dom:RARRCCTX	+ 937.00	1359.73	1.1e-67	1855	X76466 Rabbit renal cortical N
gb_pat:AR036570	+ 932.00	1350.53	3.6e-67	2266	AR036570 Sequence 18 from pate
gb_p3:HSU90544	+ 932.00	1350.47	3.6e-67	2281	U90544 Human sodium phosphate
gb_hcg2:AL1138726	+ 876.50	1230.66	1.7e-60	140266	AL1138726 Homo sapiens chromo
gb_p3:HS283953	+ 615.50	885.56	2.9e-41	2512	AJ387747 Homo sapiens mRNA for
gb_pat:140028	+ 595.00	869.35	2.3e-40	570	Z83953 H.sapiens mRNA; clone CD
gb_pat:140029	+ 479.00	684.71	4.4e-30	2716	I40028 Sequence 1 from patent
gb_pat:140220	+ 479.00	684.71	4.4e-30	2716	I40220 Sequence 3 from patent
gb_pat:140221	+ 479.00	684.71	4.4e-30	2716	I40221 Sequence 3 from patent
gb_pat:173259	+ 479.00	684.71	4.4e-30	2716	I73259 Sequence 1 from patent
gb_pat:173260	+ 475.00	684.71	4.4e-30	2716	I73260 Sequence 3 from patent
gb_ro:RNU07609	+ 475.00	681.59	6.5e-30	2024	U07609 Rattus norvegicus brain
gb_in2:AF095787	+ 466.50	668.44	3.5e-29	2178	AF095787 Caenorhabditis elegans
gb_hcg4:AC014246	+ 445.50	612.17	4.8e-26	33444	AC014246 Drosophila melanog
gb_hcg6:AC020077	+ 443.50	605.92	1.1e-25	47723	AC020077 Drosophila melanog
gb_in2:AF024691	+ 440.50	616.38	2.8e-26	9703	AF024691 Drosophila ananassa
gb_hcg4:AC014600	+ 430.50	589.96	8.3e-25	34229	AC014600 Drosophila melanog
em_in:AC003924	+ 430.50	581.73	2.4e-24	82692	AC003924 Drosophila melanog
gb_in2:AC006302	+ 430.50	576.86	4.5e-24	139238	AC006302 Drosophila melanog
gb_hcg4:AC013869	+ 419.50	583.96	1.8e-24	11568	AC013869 Drosophila melanog
gb_hcg4:AC015336	+ 417.00	585.24	1.5e-24	6806	AC015336 Drosophila melanog
gb_hcg12:AC009391	+ 415.00	549.51	1.5e-22	228783	AC009391 Drosophila melanog
gb_in2:AC004345	+ 410.00	551.26	1.2e-22	86419	AC004345 Drosophila melanog
gb_hcg5:AC018326	+ 410.00	548.97	1.6e-22	110482	AC018326 Drosophila melanog
gb_in2:AC004246	+ 407.50	552.92	1.0e-22	50279	AC004246 Drosophila melanog
gb_hcg6:AC019866	+ 407.50	548.84	1.6e-22	75607	AC019866 Drosophila melanog
gb_in1:DNMADCO7R	+ 404.50	575.63	5.2e-24	25674	I077720 D.melanogaster mRNA for
gb_hcg6:AC019757	+ 400.50	539.81	5.2e-22	66241	AC019757 Drosophila melanog
gb_hcg6:AC020165	+ 400.00	562.26	2.9e-23	5527	AC020165 Drosophila melanog
gb_hcg6:AC020123	+ 400.00	539.92	3.1e-22	60528	AC020123 Drosophila melanog
gb_hcg5:AC008232	+ 397.00	529.02	2.1e-21	121584	AC005463 Drosophila melanog
gb_hcg3:AC005463	+ 390.00	522.02	5.1e-21	85651	AC005463 Drosophila melanog
gb_in2:AC005448	+ 390.00	521.89	5.1e-21	86889	AC005448 Drosophila melanog
gb_hcg11:AC007084	+ 390.00	517.52	9.0e-21	138793	AC007084 Drosophila melanog

gb_hcg8:AC011065	+ 387.50	513.94	1.4e-20	137525	AC011065 Drosophila melano
gb_hcg6:AC020394	+ 387.00	525.21	3.4e-21	37983	AC020394 Drosophila melano
gb_in1:CEC38C10	- 381.50	518.13	8.3e-21	34193	I219153 Caenorhabditis eleg

seq_name: gb_pat:AR036571

seq documentation block:

LOCUS	AR036571	1780 bp	DNA		PAT	29-SEP-1999
DEFINITION	Sequence 19 from patent US 5872237.					
ACCESSION	AR036571					
VERSION	AR036571.1	GI:5953239				

KEYWORDS

Unknown.

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 1780)
Feder, J.Nachman, Krommal, G.Scott, Lauer, P.M., Ruddy, D.A., Thomas, W.,
Tsuchihashi, Z. and Wolff, R.K.

JOURNAL Patent: US 5872237-A 19 16-FEB-1999;
TITLE Megabase transcript map: novel sequences and antibodies thereto
FEATURES Location/Qualifiers

source 1..1780
/organism="unknown"

BASE COUNT 458 a 436 c 402 g 484 t

ORIGIN

alignment_scores:

Quality:	2061.00	Length:	401
Ratio:	5.165	Gaps:	0
Percent Similarity:	99.501	Percent Identity:	99.252

alignment_block:
US-09-391-958-1 x AR036571 ..

Align seg 1/1 to: AR036571 from: 1 to: 1780

1	MctlnValAspGluThrLeuIleProArgLysValProSerLeuGlyCysE	17
377	ATGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	426
17	TAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	34
427	TGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	476
34	lealaglnAsnValIlelelelelelelelelelelelelelelelele	50
477	TAGCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG	526
51	ThierProGlnSerGlnLeuAsnAspSerSerGlnValLeuProValAs	67
527	ACAAGCCCTCAATCCAGCTCAATGATGATGATGATGATGATGATGATG	576
67	pSerPheGlyLeuSerSerLysAlaProLysSerLeuProAlaLysSerS	84
577	CTCATTTGGTGGCTCACTAATAAGCCCAAGAGCTTCTCTCAAGTCTC	626
84	erileleuGlyGlnPheAlaIleIleIleIleIleIleIleIleIleIleI	100
627	CAATACCTGGGGGACAGTTTCAATTTGGAAAGATGGGGCCCTCCACAA	676
101	GUARGSerArgLeuGlySerIleAlaLeuSerGlyMetLeuLeuGlyCy	117
117	sPheThrAlaIleLeuIleGlyGlyPheIleSerGlnThrLeuGlyTrp	134
727	CTTTACTGCAATCTCATAGTGGCTTCAATTAAGTAACCTTGGGTGCG	776
134	ropheValPheTrpIlePheGlyGlyValGlyCysValCysLeuLeu	150
777	CCCTTGTCTCTATATATCTTGTGAGGTGTTGGCTGTGCTGCTGCTCTC	826

```

151 TrpPheValValIleTyraSpaSPProValSerTyProTrIleSerTh 167
|||||
827 TGGTTTGGTGGATTATGATGACCCCTTTCTATCCATGATGATAAGAC 876
|||||
167 rSerGluGluGluTyrIleIleSerSerIleuGluGluValIleSers 184
|||||
877 CTCGAAAAAGAAATACATCATCTCTCTGAAACAAACAGTCGGTCTT 926
|||||
184 eTyrGlnProLeuProIleLysAlaMetLeuAfsSerLeuProIleTip 200
|||||
927 CTAGACAGCCCTTCCCATCAAAAGCTATGCTCATCTCTACCATTTGG 976
|||||
201 SerIleGlySerGlyCysPheSerHisGlnTrpLeuValSerThMetVa 217
|||||
977 TCATATATGTTAGGCTGTTTCAGCCATCAATGTTAGTACACAAATGT 1026
|||||
217 ValTyTIlProThrTyIleSerSerValTyHisValAlaSerIleArg 234
|||||
1027 TGTATACATACCAACTTACATACAGCTCTGTGTACCATGTTAACATCAG 1076
|||||
234 sPaSngIleuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
|||||
1077 ACATAGACTTCTATCTGCCCTTCCTTTATGTGTGCTGGGTCAATAGC 1126
|||||
251 MetValGlyGlyTyTyrLeuAlaAspPheLeuThrlYsIysPheArgle 267
|||||
1127 ATGTGTGGAGGCTATCTGGCAGATTTCTTCTTACCAAAAAAGTTTACGT 1176
|||||
267 uIleThrValArgLysIlleAlaThrlleuGlySerLeuProSerSerA 284
|||||
1177 CATCAGCTGTAGGAAATTCGCCAATTTTACGAAAGTCTCCCTCTTCAG 1226
|||||
284 lAleuIleValSerLeuProTyTyrLeuAsnSerGlyTyTyrIleThrlAlaTh 300
|||||
1227 CACTCATTTGTCTCTGCCCTTACCTCAATTCGGGTATATACAGCAAGT 1276
|||||
301 AlaleuLeuThrlleuSerGlyGlyLeuSerThrlleuGlyGlnSerGlyI 317
|||||
1277 GCCTTCTGACGCTCTCTTGGCGATTACGACATTTGTGTACAGCAGGAT 1326
|||||
317 eTyTyrIleAsnValIleuAspIlleAlaProArgTyTyrSerSerPheLeuMetG 334
|||||
1327 TTATATCATATGCTTATGATATTGCTCCAAAGTATTCAGATTTCCTATGG 1376
|||||
334 lYAlaSerArgGlyPheSerSerIlleAlaProValIleValProThrVal 350
|||||
1377 GAGCATCAGAGAGATTTCAGACATAGCACCTGTCAATTGACCACTGTC 1426
|||||
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValPh 367
|||||
1427 ACGGATTTCTTCTTAGTACAGACCTGAGTTGGGTGAGAGAAATGTCTT 1476
|||||
367 ePheLeuLeuPheAlaValAlaSnLeuLeuGlyLeuLeuPheTyTyrIleIleP 384
|||||
1477 CTTCTCTGCTGTTTGGCGTTAACCTGTTAGACTACTCTTCACTCATAT 1526
|||||
384 heGlyGluAlaAspValAlaGlnGluTrpAlaLysGluArgLysLeuThrArg 400
|||||
1527 TTGGAGAACAGATGTCACAAAGATGGGCTTAAAGAGAGAAACCTCACTCGT 1576
|||||
401 leu 401
|||
1577 TTA 1579

```

seq_name: gb_pr3:HSU90545

seq_documentation_block:

LOCUS HSU90545 1795 bp mRNA PRI 02-MAY-1997
DEFINITION Human sodium phosphate transporter (NPT4) mRNA, complete cds.
ACCESSION U90545
VERSION U90545.1 GI:2062691
KEYWORDS human.
SOURCE human.

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1795)
AUTHORS Ruddy,D.A., Kronmal,G.S., Lee,V.K., Mintler,G.A., Quintana,L.,
Domingo,R. Jr., Meyer,N.C., Basava,A., McClelland,E., Fullan,A.,
Maps,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchinashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
A 1.1 megabase transcript map of the human hereditary
hemochromatosis locus
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1795)
AUTHORS Ruddy,D.A., Kronmal,G.S., Lee,V.K., Mintler,G.A., Quintana,L.,
Domingo,R. Jr., Meyer,N.C., Basava,A., McClelland,E., Fullan,A.,
Maps,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchinashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
Direct Submission
JOURNAL Submitted (25-FEB-1997) Sequencing, Mercator Genetics, 4040
Campbell Avenue, Menlo Park, CA 94025, USA
FEATURES
location/Qualifiers
1..1795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
377..1582
/gene="NPT4"
377..1582
/gene="NPT4"
/codon_start=1
/product="sodium phosphate transporter"
/protein_id="AAB53423.1"
/db_xref="GI:2062692"
/translation="MOVDELILPRKGSLSARYGALVLFNCFTTIAQNVINITM
VAMVNSTPQSLNDSSEVLVDSFGSLSKAPKSLIPAKSSILGGQFAIWMKGPQDR
SRLSIALSGMLGCFPAIILIGFISELIGMPFVYIEGGVGCCLMFVYIYDDFE
SYPISTSEKEYIISLKKOOGSSKOPLIKAMRSIPISICGCSHOMVSTWY
YIPYISSVYHVNIRDNGILSALPEIYAWYIGWGYGLADFLTKRFLYTRKTIAT
IGSPSSALIVSLEPLNSGTTTATLALTLSCGSTLQSGIYINVDIARYSEFLNG
ASRFSSTIAPIYTVTVSGFLISDPEFGWNVFLFPAVNLGLLFYLIIFGEADVQEM
AKERRKRL"
BASE COUNT 473 a 436 c 402 g 484 t
ORIGIN
alignment_scores:
Quality: 2061.00 Length: 401
Ratio: 5.165 Gaps: 0
Percent Similarity: 99.501 Percent Identity: 99.252
alignment_block:
US-09-391-958-1 x HSU90545 ..
Align seg 1/1 to: HSU90545 from: 1 to: 1795
1 MetGlnValAspGluThrLeuIleProArgLysValProSerLeuCysE 17
|||||
377 ATGCAAGTGGATGAGACACTGATCCCGAGAAAGTCCAAAGTTATGTTGTC 426
|||||
17 TAlaArgTyGlyIleAlaLeuValLeuHisPheCysAsnPherThrI 34
|||||
427 TGCTCGTATGGAATAGCCCTGCTTACATTTTCGCAATTTCAACAAGA 476
|||||
34 lAlaGlnAsnValIleMetAsnIleThrMetValAlaMetValAsnSer 50
|||||
477 TAGCACAATAATGTCATCATGAAACATCACCATGCTAGCCATGCTCAACAGC 526
|||||
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
|||||
527 ACAAGCCCTCAATCCACAGCTCAATGATTCCTCTGAGGTGCTGCGTTGA 576
|||||
67 pSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSers 84
|||||

```

```

577 CTCATTGTGGCCTAAGTAAAGCCCAAGAGTCTCTGCAAAAGTCC 626
84 ertleuuglyglylnphealaltrppgluatgtrpglyproproglu 100
|||||
627 CAAACTTGGGGGTCAGATTGGCAATTTGGGAAAGTGGGGCCCTCCACAA 676
101 GluArgSerArgLeuCysSerIleAlaLeuSerGlyMetLeuLeuGlyC 117
|||||
677 GAACGAAGCAGACTCTGCACATTCCTTTATCAGGAATGTTACTGGGATG 726
117 spherhAlaIleleuileglylypheIleSerGluThrLeuGlyTrp 134
|||||
727 CTTTACTGCCATCCTCATAGGTGGCTTCATTAGTGAACACCCTTGGGGCG 776
134 ropheValPheTrpIlePheGlyGlyValGlyCysValGlyCysLeuLeu 150
|||||
777 CCTTGGCTCTATATCTTGGAGGTGTGGCTGTGCTGCTGCTCTCTC 826
151 TrpPheValAlIeTrpAspAspProValSerTrpProTrpIleSerTr 167
|||||
827 TGGTTTGTGGATTATGATGACCCCTTTTCTTATCCATGAGATNAGCAC 876
167 rSerGluLysGluTrpIleIleSerSerLeuLysGlnGlnValGlySers 184
|||||
877 CTCAGAAAAGAAATACATCATATCTCTTGAACAACAGGTGGGCTCT 926
184 eTrpGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTrp 200
|||||
927 CTAGAGCGCTCTCCATCAACATATGCTGACATCTTACCCATTCTG 976
201 SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMet 217
|||||
977 TCCATATGTTTAGCTCTTTCAGCCATCATAGTTAGTACCAAAAGGT 1026
217 lValTrpIleProThrTrpIleSerSerValTrpHisValAsnIleArg 234
|||||
1027 TGTATACATACCAACTACATACAGCTGTGTACCATGTTAACTACAG 1076
234 spasnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
|||||
1077 ACAATGAGACTTATCTGCCCCCTCTTTATGTTGGCTGGGCTATGAGC 1126
251 MetValIlyGlyTrpLeuAlaAspPheLeuThrLysLysPheArgLe 267
|||||
1127 ATGTGGGAGAGCTATCTGGCAGATTTCTTCTTAAACCAAAAGTTAGCT 1176
267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSer 284
|||||
1177 CAACTGCTGAGGAAATTTGCCACAATTTTAGAAGAGTCTCCCTCTCAG 1226
284 lAlaLeuIleValSerLeuProTrpLeuAsnSerGlyTrpIleThrAlaThr 300
|||||
1227 CACTCATGTTGTCTCTGCTTACTCTCAATTCGGGCTATACACAGCACT 1276
301 AlAlaLeuThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGly 317
|||||
1277 GCCCTGTGGAGCTCTCTGGGGATTAAAGCATTGTGTACGTAGGAGAT 1326
317 eTrpIleAsnValLeuAspIleAlaProArgTrpSerSerPheLeuMetG 334
|||||
1327 TTTATATCAATGTCTTATATATGTGTCCAGGATATTTCCAGTTTCTCAT 1376
334 lValAspArgGlyPheSerSerIleAlaProValIleValProThrVal 350
|||||
1377 GAGCATCAAGAGATTTCGAGCATACACCTGTCAATGTACCACTGTCG 1426
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpParGAsnValPh 367
|||||
1427 AGCGGATTTTCTTCTTACGTACAGACCTGAGATTGTGGTGAGGAATGCT 1476
367 ePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTrpLeuIleP 384
|||||
1477 CTCTCTGCTGTTGGCGTTAAACGTGTAGGACTACTCTTCTACTCATAT 1526

```

```

384 heGlyGluAlaAspValGlnGluTrpAlaLysGluArgLysLeuThrArg 400
|||||
1527 TTGGAGAGAGGAGATGCTCAAGATGGGCTTAAAGAGAAAACTGCTGCT 1576
401 Leu 401
|||
1577 TTA 1579

seq_name: gb_prl:AB020527
seq_documentation_block:
LOCUS AB020527 2626 bp mRNA PRI 14-MAY-1999
DEFINITION Homo sapiens mRNA for Na/P04 cotransporter homolog, complete cds.
ACCESSION AB020527
VERSION AB020527.1 GI:4587206
KEYWORDS Na/P04 cotransporter homolog.
SOURCE Homo sapiens tissue_l1b:small intestine cDNA to mRNA,
clone:kaia2138.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Shibui,A., Tsunoda,T., Seki,N., Suzuki,Y., Sugane,K. and Sugano,S.
TITLE Isolation and chromosomal mapping of a novel human gene showing
homology to Na/P04 cotransporter
J. Hum. Genet. 44 (3), 190-192 (1999)
JOURNAL
MEDLINE 99253143
REFERENCE
2 (bases 1 to 2626)
AUTHORS Sugano,S. and Shibui,A.
TITLE Direct Submission
SUBMITTED (27-NOV-1998) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science the University of Tokyo,
Department of Virology; 4-6-1 Shirokanedai, Minato-ku, Tokyo
108-8639, Japan (E-mail: ssugano@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
location/Qualifiers
source
1..2626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="kaia2138"
/map="6p22.3"
/tissue_l1b="small intestine"
/note="vector pME18S"
109..1602
/codon_start=1
/product="Na/P04 cotransporter homolog"
/protein_id="BAA76663.1"
/db_xref="GI:4587207"
/translation="MSTGPDVRAVYGDIDSSDNLNVAOECGRKGCYSVRHGLALITQ
LCNFSLYTOQMLSLAIIPAMVNTAPSPQNSASTERSYDSOGYMETLKEFKAMAPA
YMSPEIOGIIILSLNYSFLAIPSGYAGJFGAKYVYGAGLFISSFTLPLAAN
AGVALLIVIRIOGIAOVNVLNGOVSIMKNAPLERSQLTTIAGSGMLSPYLLA
GSLHCOTGMPVYVFYFGICGACCCALMPFLYDDPVNHPFISAGKRRTVCSLAOD
CSFGWSLPIRAMIKSLPLMAILLVSYFCEYMLTYYLMAVPTVYISSVLQNLBDSGLS
ALPFWGSCIIILIGLADFLLSRKILRLITRKLEFATGAVFSPVYLALBWPVSSSH
SMTMFLVLSAISISECSGALVNFIDLPRTYTGFLKGLQVFAHLAGVISPAGGF
ISDSEFGGRNVFLLSAAVNSIGLVLYLLFEGADYQDMAKKEQTFTHL"
BASE COUNT 648 a 651 c 560 g 767 t
ORIGIN

alignment_scores:
Quality: 1023.50 Length: 465
Ratio: 3.270 Gaps: 2
Percent Similarity: 67.312 Percent Identity: 43.441

alignment_block:
US-09-391-958-1 x AB020527
Align seg 1/1 to: AB020527 from: 1 to: 2626

```

```

16 CysSerAlaArgTyrGlyIleAlaLeuValLeuHisPheCysAsnPheth 32
   |||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
205 TGTTCAGTCCGACATAGGGCTGGCCCTCATCTTGCAGCTCTTAATTTTTC 254
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 rThlIleAlaIAsnValIleMetAsnIleThrMetValAlaMetValA 49
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 AATTACACCCACAAATGAATGAGATGAGCATTCACATTCACGCTATGATGA 304
   snSerThrSerProGlnSerGlnLeuAsnAspSerSerGlnValLeuPro 65
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 ACAACAGACGCCCCACAGCTAGCCAGCCCATGCTCCACAGAACGGCCCTCC 354
   66 ValAspSerPheGlyIleGlyLeuSerLysAlaProLysSerLeuProAla 82
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 ACTGACTCCGAGGGCTACTGGAATGAATCTTAAGAAATTTTAAGCAAT 404
   82 sSer.....:|||||:|||||:|||||:|||||:|||||:
405 GGGCCCTGCATATGACTGAGTCTGTAATCAGGGAATCATCTCCAGCT 454
   83 .....:|||||:|||||:|||||:|||||:|||||:
455 CCTCACTATAGGCTCATTTGTGGCTCCAAATCCCACTGGCTATGAGCT 504
   83 .....:|||||:|||||:|||||:|||||:|||||:
505 GGAATATTTTGGAGCCAGATATGTTGTTGCTGGCTTGTATTATTTCCCT 554
   83 .....:|||||:|||||:|||||:|||||:|||||:
555 ATTCCGAGCCCTCTTCATTTCCACTGGCAGCTAATGCGGGAGTGGCCCTGC 604
   84 .....:|||||:|||||:|||||:|||||:|||||:
605 TCATTGTCTCCGAAATGTACAGCATAGCCAGGTTATGTAATTAACCT 654
   88 GLyGlnPheAlaIleTrpGluArgTyrPheGlyProProGlnIleuArgSerAr 104
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
655 GGCAGATATTCATTTTGGTCAATAGGGCTCCCACTGGAAAGAGATCA 704
   104 gLeuCysSerIleAlaLeuSerGlyMetLeuGlyCysPheThrAlaI 121
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
705 ACTCACCACCATGCTGATCAGGATCGATCGGGGCTCCTCATTTGTTTC 754
   121 leuLeuIleGlyGlyPheIleSerGlnThleuGlyTrpProPheValAla 137
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
755 TACTTGCTGGTGGTCTCTCTGCCAGACCATAGGATGGCTTACGCTTC 804
   138 TyrIlePheGlyIleGlyValGlyCysValGlyCysLeuLeuThrPheValA 154
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
805 TATATCTTTGGAGGATTTGGCTGTGCTTTTCTCTCTCTGCTTCTCTCT 854
   154 IleTyrAspAspProValSerTyrProTrpIleSerThrSerGluLysG 171
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
855 CATTTATGATGATCCCTGTAATCATCCCTTATCATGCTGTGGAGAAAG 904
   171 lAtyTrIleIleSerSerLeuLysGlnIleValGlySerSerLysGlnPro 187
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
905 GATACATTTGTGTTTCATTTGCTGCTCAGCAGACTGTTCACACGAGCTG 954
   188 leuProIleLysAlaMetLeuArgSerLeuProIleTrpSerIleCysLe 204
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
955 CTTCCCATTAAGGCTATGATCAATCCCTTACACACCTGGCCCATTTTACT 1004
   204 uGlyLysPheSerHisGlnTrpLeuValSerThrMetValValTyrIleP 221
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1005 CTTTATTTCTGTGATACTGCTTTTATACCATTAATGAGCGTACACAC 1054
   221 rOthTyrIleSerSerValTyrHisValAsnIleArgAspAsnLysLeu 237
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1055 CAAGCTACATCAGCTCGGTACTTCAAGCCAACTCAGAGATAGTGGATC 1104
   238 leuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 254

```

```

|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:
1105 CTGTCTCCCTTGGCTTGTGTGGATGATGATCATTTCCCTGGAGG 1154
   yTyrLeuAlaAspPheLeuLeuThrLys...PheArgLeuIleThr 270
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1155 TCTATCTGGACACTTTCTTCTCTCCAAATAATCCCTAGACTATACCA 1204
   270 aAlaArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu 286
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1205 TCAGGAATCCTTCACATCCATTTGGGGTCTCTCCATCCGATGATCCT 1254
   287 ValSerLeuProTyrIleuAsnSerGlyTyrIleThrAlaThrAlaLeu 303
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1255 GTGTCCCTGCTGGGGCAGATCCAGCCAGCATGACATGACCTTCTT 1304
   303 urhLeuSerCysGlyLeuSerThrLeuCysGlnSerGlyIleTyrIleA 320
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1305 GGTGCTGTCTTTCGCATCGACGCTTCTGTGAATCAGAGCCCTTGTA 1354
   320 snValleuAspIleAlaProArgTyrSerSerPheLeuMetGlyAlaSer 336
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1355 ACTTCTTGATATGCTCTCGGTACACGCTTCTCAAGACATATG 1404
   337 ArgGlyPheSerSerIleAlaProValIleValProThrValSerGlyPh 353
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1405 CAAGTCTTTCACACATAGCTGAGCCATCTCTCTACTGCTGGAT 1454
   353 eLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValPhePheLeu 370
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1455 TTTCATCAGTCAGAGATTCAGAGTTGTTGGAATAATGCTTCTTCT 1504
   370 eupheAlaValAsnLeuLeuGlyLeuLeuPheTyrIleIlePheGly 386
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1505 CAGCTGCTGTAACTATCGGGCTGGTTTCTACCTCATCTTGGCCA 1554
   387 AlaAspValGlnIleuTrpAlaLysGluArgLysLeuArgLeu 401
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1555 GCAGATGTGACAGACTGGGCTAAGAGCAGACATTCACCACCTC 1599
seq_name: gb_pr2:HSNAPII
seq_documentation_block:
LOCUS HSNAPII 1549 bp mRNA PRI 25-JAN-1994
DEFINITION H.sapiens mRNA for sodium-phosphate transport system 1.
ACCESSION X71355
VERSION X71355.1 GI:450531
KEYWORDS brush border membrane; NPT1 gene; renal sodium-dependent phosphate transporter; transporter.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1549)
AUTHORS Chong,S.S., Kristjansson,K., Zoghbi,H.Y. and Hughes,M.R.
Molecular cloning of the cDNA encoding a human renal sodium phosphate transport protein and its assignment to chromosome 6p21.3-p23
JOURNAL Genomics 18 (2), 355-359 (1993)
MEDLINE 94117004
REFERENCE 2 (bases 1 to 1549)
AUTHORS Chong,S.S.
Direct Submission
Submitted (02-APR-1993) S.S. Chong, Baylor College of Medicine,
Institute for Molecular Genetics, One Baylor Plaza, Houston, Texas
77030, USA
COMMENT Related sequence: M76466.
FEATURES
source
1..1549
location/Qualifiers
/organism="Homo sapiens"
/gb_xref="taxon:9606"
/chromosome="6"
/dev_stage="adult"
/clone_lib="lambdaMAX kidney-cortex cDNA (Clontech)"

```

```
gene
CDS
    /map="6p21.3-p23"
    13..1416
    /gene="NP1"
    13..1416
    /gene="NP1"
    /codon_start=1
    /product="Sodium-Phosphate Transport System 1"
    /protein_id="CAA50490.1"
    /db_xref="GI:450532"
    /db_xref="SWISS-PROT:O14916"
    /translation="MQDNRLPPKKVPFCSEFRGLSHLVHCCNVIIAORACNLMT
    VVVNSTDPHGLPNTSTKLLDINKPMYNMSPDIOGIIISTYGVIIIOVPYFYS
    GYSYKMGFALCLSVLSLIPPAAGIVAMVYCAVGAAGIVATAOFEIYVK
    WAPLERGLTMSNSRGLRGPFIYLVTVYCSLSGPMVFIYAGCANCLMFV
    LFYDDPKRPHCISISEKEYITSSLVQYSSRSQSLPIYAILKSLFVMAISGFTFEV
    SHNMTLYTPMFINSMHVNIKENGLSSLPYLFAMVIGNLGOLDFEFLRNILSVI
    AVRKLFYTAGFLPAIFGVCLPYLSIFYIYFIIAGATGSFCLGIVFINGLDIAP
    RVFGIKACSTLTGMIGLSIYTLGLIKODPESAMFKTFLMAINVTGLIFYLIV
    AFAEIODAKKEKHRL"
BASE COUNT      381 a      356 c      345 g      467 t
ORIGIN
alignment_scores:
    Quality: 964.00      Length: 467
    Ratio: 3.080      Gaps: 3
    Percent Similarity: 67.024      Percent Identity: 41.328
alignment_block:
US-09-391-958-1 x HSNAP11 ..
Align seg 1/1 to: HSNAP11 from: 1 to: 1549
1 MetGlnValaspIuThrLeuIleProAlgLysValProSerLeuCysSe 17
  |||||:||||: ||| |||||:||||:|||||
13 ATCCAATGATGATAACCGGTGCTCCCAAAAAGTTCAGATTCTGTTG 62
  |||||:||||:|||||
17 fAlaArgTyrGlyLeuAlaLeuValLeuHisPheCysAsnPheThrIrt 34
  | |||||:||||:||||| |||||
63 CTTTCCTATGATGATGCTTCTTCCCTTGACACGTGTGTAATGTTAATAA 112
  |||||:||||:|||||
34 LeuIaGlnAsnValIleMetAsnIlePheMetValAlaMetValAsnSer 50
  |||||:||||:||||| |||||:||||:|||||
113 CACGACAGCGTGGCTGCCTCAACCTCAATGATGATGATGATGATGATG 162
  |||||:||||:|||||
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGlnValLeu..... 64
  |||||:||||:||||| |||||
163 ACAGATCCACATGATGTTGCCACACACCTCCACAAAGAAAGCTCCTGGAT 212
  |||||:||||:|||||
64 ..... 64
213 TATAAAGAACCTATGATTAATTGAGACCCAGATATCCAGGAAATCATCT 262
  .....
65 ..... ProValAspSer 68
  |||||
263 TGAATCCACCTCCTATAGTGTGTCATCATCCAAAGTTCCTGTTGGATAC 312
  |||||
69 PheGlyGlyLeuSerLysAlaProLysSerLeu..... 79
  |||||:||||:||||| |||||
313 TTCTCTGATATATATCTACAAAGAAATGATGCTTGCATTATATCCT 362
  .....
79 ..... 79
363 CAGCTCTGTGTAAAGCTGCTCATCCACAGCAGCTGGAATGGAGTAG 412
  .....
80 ..... ProAlaLysSerSerIle 85
  |||||:||||:|||||
413 CTGGGCGCTGTGATGTCGACAGATTCAAGGAGAGCCAGGGGATAGTT 462
  .....
86 LeuGlyGlyLeuPheAlaIleTyrGlyLysTyrPheProGlnGluIrt 102
  |||||:||||:||||| |||||
463 GCAACAGCCAGTTTGAATATATATGTCAAATGAGGCTCCTCCCTCGAAGC 512
```

```
102 gSerArgLeuCysSerIleAlaLeuSerGlyMetLeuGlyCysPhe 119
  |||||:||||:||||| |||||:||||:|||||
513 AGCCGACTTACTTATGATGATACATGAGGTTTTCCTGGACCTTTA 562
  .....
119 hAlaIleLeuIleGlyPheIleSerGlnThrLeuGlyTyrProPhe 135
  |||||:||||:||||| |||||:||||:|||||
563 TTGACCTACTGTCATGAGATGATATCTGTGAATCTGAGGCTGACCATG 612
  .....
136 ValPheTyrIlePheGlyGlyValGlyCysValCysLeuLeuTyrPhe 152
  |||||:||||:||||| |||||:||||:|||||
613 GTCTCTATATATTTTGGTGTGCTGAGGCTGACCTATGCTCTCTGTT 662
  .....
152 eValValIleTyrAspAspProValSerTyrProTyrIleSerThrSerG 169
  |||||:||||:||||| |||||:||||:|||||
663 CGTTCTTTATATGATATACCCCAAGACCCCTGATTAAGCATCAGTGTG 712
  .....
169 LuYsgIuTyrIleIleSerSerLeuLysGlnGlnValGlySerSerLys 185
  |||||:||||:||||| |||||:||||:|||||
713 AAAAGAAATACATACATCCCTCCCTGTCACAGCAGTCAGTCAAGTAGA 762
  .....
186 GlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTyrSerI 202
  |||||:||||:||||| |||||:||||:|||||
763 CAATCTCTGCTATCAAGCTATCTTAAGTCCCTCAGTCTGAGCTAT 812
  .....
202 eCysLeuGlyCysPheSerHisGlnTyrPheValSerThrMetValTyr 219
  |||||:||||:||||| |||||:||||:|||||
813 TTCCATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 862
  .....
219 TyrIleProThrTyrIleSerSerValTyrHisValAsnIleArgAspAsn 235
  |||||:||||:||||| |||||:||||:|||||
863 ACATCCAAATGTTATATCAACCTCAGCTGCTCATGTTATATATAAGAAAT 912
  .....
236 GlyLeuLeuSerAlaLeuProPheIleValAlaTyrValIleGlyMetVa 252
  |||||:||||:||||| |||||:||||:|||||
913 GGGTCTTCTGCTTCCCTCATTTGTTGCTGAGTCTGATGATGATGAT 962
  .....
252 IGLGlyTyrLeuAlaAspPheLeuThrLys... PheArgLeuI 268
  |||||:||||:||||| |||||:||||:|||||
963 AGCAGGTACATGATACAGATCTTCTGACCAAGAAATTTCTCAGCCTAA 1012
  .....
268 IeThrValArgLysIleAlaThrIleLeuGlySerIleProSerSerAla 284
  |||||:||||:||||| |||||:||||:|||||
1013 TTGCTGTCGCGAATCTTTCACACACAGAGATTTCTCCTCCGCAATC 1062
  .....
285 LeuIleValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAl 301
  |||||:||||:||||| |||||:||||:|||||
1063 TTGCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1112
  .....
301 aLeuLeuThrLeuSerGlyLeuSerThrLeuGlyGlnSerGlyIleT 318
  |||||:||||:||||| |||||:||||:|||||
1113 TTTCCTATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1162
  .....
318 TyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetGly 334
  |||||:||||:||||| |||||:||||:|||||
1163 TTTAATAATGCTTGTGATATTCCTCCAGATATTTTGGATTTATATAAGCA 1212
  .....
335 AlaSerArgGlyPheSerSerIleAlaProValIleValProThrValSe 351
  |||||:||||:||||| |||||:||||:|||||
1213 TGTCTCACTTAACTGGAATGATGAGAGACTAATTTCTTCCACTTTGAC 1262
  .....
351 rGlyPheLeuLeuSerGlnAspProGluPheGlyTyrPheArgAsnValPhe 366
  |||||:||||:||||| |||||:||||:|||||
1263 TGGATTTATCTCTTAAGCAGATCCGGAATCCGCTGCTGTTAAACCTTCA 1312
  .....
368 hLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuIlePhe 384
  |||||:||||:||||| |||||:||||:|||||
1313 TCTGATGAGCAGCATTAATGTCAGTCCCAATTTTCTTACTTATATAGTT 1362
  .....
385 GlyIuAlaAspValGlnGluTyrPheAlaLysGluArgLysLeuThrArgLe 401
  |||||:||||:||||| |||||:||||:|||||
1363 GCTACACAGCAATTCAGAGCTGGGCTTAAGAAACAAACACACACAGCTCT 1412
  .....
401 u 401
```


[illegible]

seq_name: gb_ro:MMNP1CT

seq. documentation block:	
LOCUS	MNPNP1.CT 1885 bp mRNA
DEFINITION	M.musculus Npt1 mRNA for sodium-dependent phosphate cotransporter
ACCESSION	X77241
VERSION	X77241.1 GI:887522
KEYWORDS	NPT1 gene; sodium-dependent phosphate cotransporter.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	1 (bases 1 to 1885)
AUTHORS	Chong, S. S., Kozak, C. A., Liu, L., Bordeau, J. E., Hughes, M. R.,

TITLE	JOURNAL	REFERENCE
Cloning, genetic mapping and expression analysis of a mouse renal sodium-dependent phosphate cotransporter	Am. J. Physiol. 268, 1038-1045 (1995)	2 (bases 1 to 1885)

TITLE Direct Submission
JOURNAL Submitted (21-JAN-1994) S.S. Chong, Baylor College of Medicine,

FEATURES

Location/Qualifiers

```

source
1..1885
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="13"
/dev_stage="adult"
/tissue_type="kidney"
/clone_lib="1.4 kb cDNA in pSPORT1 (BRL) plasmid vector"
/map="close to tcr3"
80..1477
/gene="Npl1"
80..1477
CDS

```

```

/gene: nplc1
/codon_start: 1
/product: "sodium-phosphate transport system 1"
/protein_id: "CAA54459.1"
/db_xref="GI:887523"
/db_xref="MGD:MGI:103209"
/db_xref="SMTS-PROT:Q01983"
/translation: "MEHQCLPKKVPSCFRRGLATLHFCNIALMAQRVCLNTMVA
VAVNTGSPHLSNLSVYVSNLPPYSSNPIDIGLLISVYFGVAVVQAPVGLSG"

```

BASE COUNT	474 a	438 c	412 g	561 t
ORIGIN	<p> YPMKRITGGSSFLFSLMSLLIIPPAOAGALVIYCRILQIAGCTVSTGCHETWYKKA PPEDEKRLISTMTLSGEVWGPETVLLVSGFICDLDGMPWVFTHIGVCCVLSWFFLE FDDEKRDHPYMSSEKDYIISLMQOASSGRSLPIKMLKSLPMAIILNSFAFIWNSN SLATVYPTPTFSIVLHYNVRENGSLSLPYLLAYICGILAGMSDFLTRIFESIVTV RLKPTTSGPCPVIFIMCLOLYSYNFSTVFELANSTLSFSYCGOLINADLAPPR YGFELKAVTALIGMRGGLISSTLGLLINOPEYAMHHSFLMGINVTCLVFETFLPAK GEIOLDMAKEIKTTRLL" </p>			
alignment_scores:				
Quality:	951.00	Length:		466
Ratio:	3.009	Gaps:		3
Percent Similarity:	67.811	Percent Identity:	40.343	
alignment_block:				
US-09-391-958-1 x MMNP1ICT	..			
Align seg 1/1	to: MMNP1ICT	from: 1	to: 1885	

2 GlnValAspGluThrLeuIleProArgLysValProSerLeucysSerAl 18
77 CAAAGGGAGAACCCATGCCCTCCCAGAAGAAAGTCCAGGGTTCTGTTCCTT 126
18 ArgTrpCylLeuAlaLeuValLeuHisPheCysAsnPhetThrIleA 35
127 CCGGATGAGACTGGCTATACCTCTTGCACTCTTGTAACATTGCCATATGG 176
35 IagIAsnValIleMetAsnIleThrMetValAlaMetValAsnSerThr 51
177 CCCAGAGAGTGGTGCCTGAACCTCACCATGTGATGCCATGTGTGAACAACACA 226
52 SerProGlnSerGlnLeuAsnAspSer..... 60
227 GGGTCACCCCCCATTTATCTAACCAACTCTGATGAGAGATGTGGACAAATG 276
60 60
277 AAGAACCCCTGTGATTCTCTGGAGTGCATGATTCAGAGGGCTTATTCTCA 326
60 60
327 GTTGTGCTCTTCTTGGTATGATGTCGTGTTCAAGGCTCTCGGGGTACCTA 376
61 SerGluValIleuProValAspSerPheGlyGlyLeuSerLysAlaProLy 77
377 TCTGGAAATAATACCACATGAAGAAATAAATTGGGCTTTCACCTGTTCTCTCAG 426
77 sSerLeuPro..... 80
427 CTCCTTGGATGTCCTTGCTGCATCTCCACACTCTGCACAAATCGGAGCTGCTT 476
81 AlasSerSerIleLeu 86
477 TAGTCATGTGTGTCGAGNATCCAGAGGAATAGCCAGGGAACAGTGTCA 526
87 GlyIleGlnPheAlaIleTrpGluArgTrpGlyProProGlnGluArgSe 103
527 ACAGAGCCAGCATGAATATGGCTCAAAATGGGCAACCTCCCTTAGAACGAGG 576
103 PArgLeuCysSerIleAlaLeuSerGlyMetLeuLeuGlyCysPheThrA 120
577 CAGACTTACCTTAAGACCTTAACAGGGTTGTAAAGGAGCCATTATATG 626
120 IaIleLeuIleGlyIlePheIleSerGluThrLeuGlyTrpPropheval 136
627 TCCTGCTGTGAGTGGCTTCATCTGTGATCTTCTGTGGCTGGCCCATGTGC 676
137 PheTrpIlePheGlyGlyValGlyCysValCysCysLeuLeuTrpPheVa 153
677 TTTCACATCTTTGGTATGTGGGGTGTGTGTGAGAGTCTTCCCGTGTCTT 726

```

153 lvalileryaspaspProvalSerTyrrProtripleSerThrSerLut 170
      ::::::::::::::::::::|
727 TCATCTTGTATGATGACCCAGACCCGCTATATAGACAGCAGTAGA 776
170 yslgufyrllelleSerSerLeuLysGlnGlnValGlySerSerLysGln 186
      |::::::::::::::::::|
777 AGACATACATATATCTCCCTCATGACACAGCCAGCTCAGCAGACAA 826
187 ProLeuProLleLysAlaMetLeuArgSerLeuProLleSerLecy 203
      |::::::::::::::::::|
827 TCCTGCGCATCAAGCATATGCTTAATCTCTCCACTGCGCCATTAT 876
203 sleuglyCysPheSerHisGlnTrpLeuValSerThrMetValValTyr 220
      |:::::|
877 CCTCAATAGCTTGTTCATATGTCGACAGACGTCCTCGTTACATACA 926
220 leprothryrileSerSerValTyrHisValAsnleatgaspasnly 236
      |::::::::::::::::::|
927 CGCGACATTATATACACACTGCTCATGTTAATGTTAGAGAAATGGA 976
237 LeuLeuSerAlaLeuProPheileValAlaTrpValIleGlyMetValGI 253
      |::::::::::::::::::|
977 CTGCTCTCCAGCCTCCCTATCTGCTTGCCTACATCTGTGTATCCTAG 1026
253 yglYrYrLeuAlaAspPheLeuLeuThryrLys...PheArgLeuIleT 269
      |:::|
1027 AGGTGAGATGTACAGACTCTCTCTGACACAGAAATTTTCAGCATAGT 1076
269 hvValArgYrYrIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu 285
      |::::::::::::::::::|
1077 CGGTCCGGAACCTCTACACACTAGGAGATTTGTGCCGATCTTC 1126
286 lIleValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAla 302
      |:::::|
1127 AHCATGTGCTGCTACTGACGTACAACTTCTACAGACACGCTCTTTT 1176
302 uLeuThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGlyIleTyr 319
      |:::::|
1177 CTGACACTTGGCACTCAACACTTACTTCTCTCTGACGCAACTCA 1226
319 lAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetGlyAla 335
      |::::::::::::::::::|
1227 TCATGCTGTGATGATGCTCCAGATATATGATTTCTTAAGGACAGTC 1276
336 SerArgGlyPheSerSerIleAlaProValIleValProThrValSerGI 352
      |:::::|
1277 ACAGCTTAAATGGAATGTTGGGGCCGATTTCTTCAACTCTGCTGG 1326
352 yPheLeuLeuSerGlnAspProGluPheGlyTyrTrpArgAsnValPhePhe 369
      |:::::|
1327 GTTGATTTCTTAATCAGATCCGGAATATGCTTGATTAATAATTTCTTCT 1376
369 euleuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuIlePheGly 385
      |:::::|
1377 TGATGCGACGATTAATGATGACATGCTGCTTCTAATTTCTAATTTGCT 1426
386 GluAlaAspValGlnGluTrpAlaLysGluArgLysLeuThryrArgLeu 401
      |::::::::::::::::::|
1427 AAAGAGAAATTCAGGACTGGCTTAAGAAATTAATAAACACACGACTG 1474
      |::::::::::::::::::|
seq_name: gb_Prl.HUMAPI4
seq_documentation_block:
LOCUS HUMAPI4 1794 bp mRNA Pri 07-FEB-1999
DEFINITION Human mRNA for renal Na+-dependent phosphate cotransporter,
complete cds.
ACCESSION D28532.1 GI:639841
VERSION D28532.1
KEYWORDS Na/Pi-4; renal Na+-dependent phosphate cotransporter.
SOURCE Homo sapiens adult male kidney cortex cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.

```

```

REFERENCE 1 (bases 1 to 1794)
AUTHORS Miyamoto,K.-I.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1994) to the DDBJ/EMBL/GenBank databases.
101 CACAGCTATGCTGACCTGACCTCCACAAATGATGATGATGATGATGACCA 150
151 GATCCACATGTTGGCCACACACCTCCACAAAGAAAGCTCTTGATTAATAT 200
64 .....
201 AAAGAACCTATGATTAATGATGAGCCAGATGTCAGGAAATCATCTTGA 250

REFERENCE 2 (bases 1 to 1794)
AUTHORS Miyamoto,K., Tatsumi,S., Sonoda,T., Yamamoto,H., Minami,H.,
Taketani,Y. and Takeda,E.
TITLE Cloning and functional expression of a Na(+)-dependent phosphate
co-transporter from human kidney: cDNA cloning and functional
expression
JOURNAL Biochem. J. 305 (Pt 1), 81-85 (1995)
MEDLINE 95126933
FEATURES
source Location/Qualifiers
1..1794
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="adult"
/sex="male"
/tissue_type="kidney cortex"
4..1401
/gene="Na/Pi-4"
4..1401
/gene="Na/Pi-4"
/codon_start=1
/product="renal Na+-dependent phosphate cotransporter"
/protein_id="BA05888.1"
/db_xref="GI:639842"
/translation="MDNRDLPKKRPPKRCSPRGISFLVHCCNVTITRAQLMCLNTMNV
MVNSTDPHGIPNSTKRLKLDIKPMNMSPDVGITLSTSYGVIIIOVGVFSGI
YSTKMGFALCLSSVSLIPLPAAGVAMVVCRAVQAGIYVTAQFEIYKWA
PPLERGRILSMSTSGFLGPEIYLVLTGIVGAMVYIFGACCAVCLMEYVF
YDPRKDCISIGEKYITSLVOVSSRSOSLPKAILSLPVMATISGTFEFSH
NIMTLVPMFPLNSMLHYNIKENGFLSLPLFMICGNLAGOLSDPTITNITSVIAY
KILFTAGFLLPATEFGVCLPYLSTSIYIFILACATGSCFGLGVFIIGLDIAPY
FGRIKASTITMGIGLITASTLTGLIKDPESAWKFTILMAINVTGLIFLIYAT
AETIDWAKEKHQHTL"
BASE COUNT 460 a 398 c 385 g 551 t
ORIGIN
alignment_scores:
Quality: 944.00 Length: 466
Ratio: 3.045 Gaps: 3
Percent Similarity: 66.524 Percent identity: 40.773
alignment_block:
US-09-391-958-1 x HUMAPI4
Align seg 1/1 to: HUMAPI4 from: 1 to: 1794
2 GlnValAspGluThrLeuIleProArgLysValProSerLeuCysSerAl 18
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
1 CAAATGGAATAACCGGTGCTGCCCAAAAAAGTTCACAGGTTCTGTCTT 50
18 argYrGlyIleAlaLeuValLeuHisPheCysAsnPheThrTrpIleA 35
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
51 CCGCATGATGATGCTCTTCTGTCACACTGTTGATGATTAATAACAG 100
35 lAsnValValIleMetAsnIleThrMetValAlaMetValAsnSerThr 51
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
101 CACAGCTATGCTGACCTGACCTCCACAAATGATGATGATGATGATGACCA 150
52 SerProGlnSerGlnLeuAsnAspSerSerGluValLeu..... 64
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
151 GATCCACATGTTGGCCACACACCTCCACAAAGAAAGCTCTTGATTAATAT 200
64 .....
201 AAAGAACCTATGATTAATGATGAGCCAGATGTCAGGAAATCATCTTGA 250

```

```

65 .....ProValAspSerPhe 69
251 GTTCAACCTCTATGGTGCATCATCATCAAGTCTCTGTGGATATCTC 300
70 G1YGLYLeuSerLysAlaProLysSerLeu.....79
301 TCTGGATATATTTCTACAAAGAAATGATTGGCTTTCATTATAGCCCTACG 350
79 .....79
351 CTTCTGTATTAGCCTGCTCATCCACAGACGCTGCAATTGGAGTACCTT 400
80 .....ProAlaLysSerSerLeu 86
401 GGGTCGTGTATGTCAGACGATTCCAGGAGACGCCAGGGGAGATTGCA 450
87 G1YGLYGLNPhelAlaIleTPG1uATGTPG1YProProGLNGLuArgSe 103
451 ACAGCCGAGTTGAAATATATGTCAAATGGGCTCTCCCTGGAGGAGAG 500
103 IArgLeuCySerLLeAlaLeuSerGlyMetLeuGLYCySpherhA 120
501 CCACCTTACTTCTATGAGTACATCAGGGGTTTTGCTGGGACCTTATTTT 550
120 IAlleuLeuIleGLYGLYPhelSerG1uThrLeuGLYTPProPhelVal 136
551 TCTTACTTGTGACTGAGATTATCTGTCAATCTCTGGCTGCCCCATGCTC 600
137 PheTYrIlePheGLYGLYValGLYCySValCySLeuLeuThrPheVal 153
601 TTTCTATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 650
153 ValIleLeuAspAspProValSerTYrProTPHLeSerThSerLul 170
651 TCTGTTTATGATGACCCCAAGACCCACCACTGATATAGATGATGATGAT 700
170 YSG1uTYrIleIleSerSerLeuLysGLNGLNValGLYSerSerLysGLN 186
701 AGGAATATACATCATCTCTCCCTGCTCAGACGATGATTCACAGTACAA 750
187 ProLeuProIleLysAlaMetLeuArgSerLeuProIleTPSerLecy 203
751 TCTCTGCTCTCAAGCTTATCTTAAGTCCCTTCCACTGCGCTATTTC 800
203 s1eugLYCyPheSerHisGLNThrPheValSerThMetValTYrI 220
801 CATTTGAGTTTACGTTTTTCTGGTCACATTAACATCATGACATATACA 850
220 IeProThTYrIleSerSerValTYrHisValAsnIleArgAspAsnGLY 236
851 CTCGCAATGTTTATCACTCCATGCTTCTATTAATATATAAGCAATGGC 900
237 LeuLeuSerAlaLeuProPheIleValAlaTYrValIleGLYMetValGL 253
901 TTTCTGTCTCTCCCTTCCCTATTTGTTGCTGATCTGTGTTAACCTACG 950
253 YGLYTYrLeuAlaAspPheLeuLeuThrLysLys...PheArgLeuIleT 269
951 AGGTCACTTATACAGACTTCTTCCGACAGCAAGAAATTTCTCAGCCTAATT 1000
269 hYrValArgLysIleAlaThrIleLeuGLYSerLeuProSerSerAlaLeu 285
1001 CTGTCCGGAACCTTTCACACAGCAGCAAGATTTCTCTTCCCTGCAATCTT 1050
286 IleValSerLeuProTYrLeuAsnSerGLYTYrIleThrAlaThrAlaLe 302
1051 GGATGCTGCTGCTGCTTACTGAGTTCACCTTACAGCAATGTCATTTT 1100
302 uLeuThrLeuSerCySGLYLeuSerThrLeuCySGLNGLYIleTYrI 319
1101 CTTAATAATCTTGTGTCACACAGCAGCTTTTGTGTTGGTGGATGTTT 1150
319 IeAsnValLeuAspIleAlaProArgTYrSerSerPheLeuMetGLYAla 335

```

```

1151 TAAATGGCTTGATATGCTCCAGATATTTTGGATTTATTAAGCATGT 1200
336 SerArgGLYPheSerSerIleAlaProValIleValProThrValSerGL 352
1201 TCAACTTAAGTGAATGATAGAGAGACTTAATCTTCCACTTGTGACTGG 1250
352 YPheLeuLeuSerGLNAspProGLupPheGLYTPArgAsnValPhePheL 369
1251 ATTGATCTTAAAGCAGATCCGAGATCCGCTGGTTTAAACCTTCATCC 1300
369 euleuPheAlaValAsnLeuLeuGLYLeuLeuPheTYrLeuIlePheGLY 385
1301 TGATGGCAGCATTAATGATGACTGCGCTAATTTCTACCTTATAGTTGCT 1350
386 GluAlaAspValGLNGLUTPAlaLysGLYArgLysLeuThrArgLeu 401
1351 ACAGCAGCAAAATTCAGAGCTGCGCTAAAGAAACACACACACGCTCTC 1398

```

seq_name: gb_com:RABRCCTX

```

seq_documentation_block:
LOCUS      RABRCCTX      1855 bp      mRNA      27-APR-1993
DEFINITION Rabbt renal cortical Na/P-1-cotransporter, complete cds.
ACCESSION  M76466
VERSION    M76466.1 GI:165689
KEYWORDS   membrane transport protein; renal cortical Na/P-1-cotransporter;
SOURCE     transmembrane protein.
ORGANISM   Oryctolagus cuniculus kidney cortex cDNA to mRNA.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE  1 (bases 1 to 1855)
AUTHORS   Werner,A., Moore,M.L., Mantei,N., Biber,J., Semenza,G. and Murer,H.
TITLE     Cloning and expression of cDNA for a Na/Pi cotransport systems of
            kidney cortex
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 9608-9612 (1991)
MEDLINE   92052140
FEATURES   source
            source
            1..1855      Location/Qualifiers
                        /organism="Oryctolagus cuniculus"
                        /db_xref="taxon:9986"
                        /tissue_type="kidney cortex"
                        /note="expression in Xenopus laevis oocytes"
                        /codon_start=1
                        /product="renal cortical Na/P-1-cotransporter"
                        /protein_id="AAA31461.1"
                        /db_xref="GI:165690"
                        /translation="MDNOPSRKGRFCFSFRVYALFHFNCNIVIIAQMCLSTINVA
                        VNNTNMLHSPNTSAERLDNTRKNPYNNSPVQGITFSIYGAFLIDIPVGIISGI
                        YSIKRLGFALFLSLVSIPIQAAVAGETWIVICVVOGIVTGTVOHBIWYKWA
                        PDLEGRRLMSVLSGLGPFIVLVATVGIICSLGPMVFIYFGACGACVCLIMFVLY
                        YDDPDHPCVSLHEKEIYRSSLIOGSSYRSLPIKAMKSLPIAMISPCFAMLY
                        SRLIYTPILINSMLHVDIIRENGLSLPIYLAAMICGVIAGHTADPFMRNMLSTAI
                        KLFTRIGLLPIVFSMCLTYSIGSFYITITLLIANNASSYSCDGGALNALDLAPRY
                        YVIRGVTTLIGMTGMSSTVAGLFLSDPSSSWPKIFLLMSIINVISIYFLIFAK
                        ABIDMAKKEKQTRL"

```

CDS

```

BASE COUNT  459 a  459 c  393 g  544 t
ORIGIN
1          polyA_signal
459 a      459 c      393 g      544 t

```

alignment_scores:

```

Quality: 937.00      Length: 466
Ratio: 3.023      Gaps: 3
Percent Similarity: 66.524      Percent Identity: 39.485

```

alignment_block:

US-09-391-958-1 x RABRCCTX ..

Align seg 1/1 to: RABRCCTX from: 1 to: 1855


```

1340 ATCAGAGTAAAGTGAATTCCTGCTCCCTCCCTTATATTCGCTGCAAG 1389
248 IILGELYEVALGILGYTYRLEUALAASPHELEULEUTHIRLYS. 264
      ::::::::::::::::::::
1390 CTGTCAATTTTAAAGAGTCAGCTGGCAGATTTCTTTGTCACAGAAATC 1439
265 ..PHEATGLEULIEHVALARGYSLLEALATHIRLEULEUCYLSERIEU 280
      ::::::::::::::::::::
1440 TTCCTCAGTTGATCAGTCGGAACCTCTTTCATCTCTTGA..... 1482
281 ProSerSerAlaLeuIleValSerLeuProTYLeuAAsnSergLYTYIIL 297
1482 ..... 1482
297 eThraAlaThraLeuLeuThraLeuSercysglyLeuSerThraLeucysG 314
1482 ..... 1482
314 InSergLYLeTYRILeAsnValLeuAspILeAlaProArgTYRISerSer 330
1483 .....TATCAAGT 1491
331 PheLeuMetGLYAlaSerARGLYPheSerSerILeAlaProVALIleVal 347
      |||||
1492 TTCCTCAGTGGGAATCTCAGGGGATTTGGGCTTCATCGCAGGAATCATCTC 1541
347 IProThraValSergLYPheLeuLeuSergInAspProGluPhegLYTPA 364
      |||
1542 TTCACATGCCACACTGATTTCTCATAGTCAGATTTTGAGTCGTGGTGA 1591
364 rGAsnValIphePheLeuLeuPheAlaValAsnLeuGLYLeuLeuPhe 380
      |||||
1592 GGAATGCTTTTCTCTGCTGCTGACAGTCACATGTTTGCTGCTGCTCTT 1641
381 TYRLeuILePhegLYGlyAlaAspValGlnGluTrpAlaIleGlyArgLY 397
      |||||
1642 TACCTCAGCTTTGGACAAGACACTTCAGACTGGGCCAAGAAGAGAAC 1691
397 sLeuThraArgLeu 401
      |||||
1692 CCTTACCGGCTC 1704
seq_name: gb_pr3:HSU90544
seq_documentation_block: 2281 bp mRNA PRI 02-MAY-1997
LOCUS HSU90544
DEFINITION Human sodium phosphate transporter (NP73) mRNA, complete cds
ACCESSION U90544
VERSION 090544.1 GI:2062689
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2281)
Ruddy,D.A., Kronmal,G.S., Lee,V.K., Muntler,G.A., Quintana,L.
Domingo,R. Jr., Meyer,N.C., Basava,A., McLelland,E., Fullan,
Mapa,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchishashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
A 1.1 megabase transcript map of the human hereditary
hemochromatosis locus
Unpublished
2 (bases 1 to 2281)
Ruddy,D.A., Kronmal,G.S., Lee,V.K., Muntler,G.A., Quintana,L.
Domingo,R. Jr., Meyer,N.C., Basava,A., McLelland,E., Fullan,
Mapa,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchishashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
Submitted (25-FEB-1997) Sequencing, Mercator Genetics, 4040
Campbell Avenue, Menlo Park, CA 94025, USA
FEATURES
SOURCE
1..2281
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
419..1729
/gene="NPT3"
419..1729
/gene="NPT3"
/codon_start=1
/product="sodium phosphate transporter"
/db_xref="GI:2062690"
/translation="MDKPRATRKGPDFCSLRGLALIMHFSNFTMITORVLSIALIA
MNTTQOGLSNASTGEVADAFNNSSISIKEDPKASYQWSPETGIISSINYG
ILTLIPSGTLAGIFGAKMGLISLILFTPLADFGVILIVMRTVOGMOG
AMTGOFTIMAKWAPPLERSKLTITLIGSAGSFIIICVGLISOLSMPTFI
TGCVCCLMEFTIYDDPMHPCISYREKEHILSSAQCPSSGRAPVIRKAVTCLPLM
AIFGFSPHFMTCTITITLPTVISTLHVNRDGSVLSLDFIAAECTILGGLAD
PLISRLRLITVRKLFSSLDKQVSWESQDLSQSSSLPLPLDSSSVRLISLVGG
MSFCLLOSTCLAMSFTRLDKQNKFTGPKRQPLPASHDIKIQT"
BASE COUNT      624 a      506 c      490 g      661 t
ORIGIN

```

```

alignment_scores:
  Quality: 932.00      Length: 471
  Ratio: 3.305        Gaps: 4
  Percent Similarity: 59.873      Percent Identity: 41.826

```

```

alignment_block:
US-09-391-958-1 x HSU90544 ..

```

```

Align seg 1/1 to: HSU90544 from: 1 to: 2281

```

```

10 ArgLysValProSerLeuCySerAlaArgTyrGlyIleAlaLeuValle 26
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 AGGAAAGGCCAGATTTCTGTTATACGTATGAGGCTGCTTTATACAT 489
26 uHisPheCysAsnPheThrThrIleAlaGlnAsnValIleMetAsnIleT 43
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 GCACCTCTCAAACTTCACCATGATTAACGACGCTGTGACGTGACGATGG 539
43 hMetValAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAp 59
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 CCATCATGCCCATGTGGAACACACCTCACGACAGCAAGCTATCTTAATGCC 589
60 SerSerGluValLeuProValAspSerPheGlyGlyLeuSerLysAlaPr 76
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 TCCACTGAGGGGCGCTGTGTCAGATGCTTCATTAATCCTCAGCATATCCAT 639
76 oLysSerLeuProAlaLysSerSerIle..... 85
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 CAAGGAAATTGATACAAAGCGCTGTGTATCATGATGAGCCCAAGAACTC 689
85 ..... 85
690 AGGGATATCATCTTTAGCTTCATCAACTATGAGATTAATACTGACTGTGATC 739
86 ..... 87
740 CCAGTGTATATTAGCAGGATATTGAGCAAAAAAATGCTTGTGTC 789
87 ..... 87
790 TGGTTGCTGATCTCTCTTCACCCCTTTTACACCACCTGCTGCTG 839
87 ..... 87
840 ACTTCGAGTATTTTGGTCATCATGGTTCGACAGCTCCAGGCGATGCC 889
88 ..... 88
890 CAGGGAATGCATGAGCAGTCACTTACTATTGGGCAAGTGGGCTCC 939

```

```

98 oProGlnGluArgSerArgLeuCySerIleAlaLeuSerGlyMetLeu 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
940 TCCACTTGAAGCAAGCAAGCTCACACCATTCAGAGATCAGGGTCAAGCAT 989
115 euGlyCysPheThrAlaIleLeuIleGlyIlePheIleSerGlyThrLeu 131
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
990 TTGGATCTCTCATCATCTCTGTGTGGGGAGCTAATCTCAGAGGCTTG 1039
132 GlyThrProPheValPheThrIlePheGlyIleValGlyCysValLysC 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1040 AGCTGGCCTTTTATCTCTCATCTTTGTTGACACGTGCTGTCTGCTG 1089
148 sLeuLeuThrPheValValIleTyrAspAspProValSerTyrProTrp 165
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1090 TCTCCATAGGTTCAAGTGTATATGATGACCCCATGATCATCACCCTGCA 1139
165 lSerThrSerGluLysGluTyrIleIleSerSerLeuLysGlnVal 181
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1140 TAAGTGTAGGAAAAGGACACATCTCTGCTCCTGCTCCTCAACAGCC 1189
182 GlySerSerLysGlnProLeuProIleLysAlaMetLeuArgSerLeuPr 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1190 AGTTCTCTGACGACGCTGCCCATTAAGGCGATGTCATGCTTACC 1239
198 oLietrPserIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerT 215
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1240 ACTTTGGCCATTTCCTGCTGTTTTCAGCCATTCTGTTATGACACA 1289
215 hMetValValTyrIleProThrTyrIleSerSerValTyrHisValAsn 231
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1290 TCATCTTAACATACCTACCAAGCTATATCATGCTGCTCATGTTAAC 1339
232 lLeuArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpVa 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1340 ATCAGAGATAGTGAAGTCTGCTCCCTGCTGCTTATGCTGCGCAAG 1389
248 lIleGlyMetValGlyGlyTyrLeuAlaAspPheLeuThrLysLys 264
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1390 CTTGTACAAATTTAGGAGGTCACCTGCGACATTCCTTTTGTCCAGAAATC 1439
265 ..PheArgLeuIleThrValArgLysIleAlaThrIleLeuLysSerLeu 280
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1440 TTCCTAGATTGATCACTGTGCGAAGCTCTTTCACTCTTCA..... 1482
281 ProSerSerAlaLeuIleValSerLeuProTyrLeuAsnSerGlyTyrI 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 ..... 1482
297 eThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuCysG 314
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 ..... 1482
314 lnsSerGlyIleTyrIleAsnValLeuAspIleAlaProArgTyrSerSer 330
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1483 ..... 1483
331 PheLeuMetGlyAlaSerArgLysPheSerSerIleAlaProValIleVa 347
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1492 TTCTCATGGAATCTCAAGGGAATTTGGCTCATGCGAAGAAATCAATCTC 1541
347 lProThrValSerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpA 364
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1542 TTCACATGCCACTGGATTCCTCATCAAGCAGAGATTTGAGCTGTGTTGA 1591
364 rGAsnValPhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPhe 380
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1592 GGAATGCTCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 1641
381 TyrLeuIlePheGlyGlyAlaAspValGlnGluTrpAlaLysGluArgLys 397
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1642 TACCTCACTGTTTGGACAGCAGAACTTCAAGACTGCGCAAGAGAGAGAC 1691
397 sLeuThrArgLeu 401

```


seq_name: gb_pr3:HSA387747	seq_documentation_block:	PRI	10-DEC-1999
LOCUS HSA387747 2512 bp mRNA			
DEFINITION Homo sapiens mRNA for statin.			
ACCESSION AJ387747			
VERSION AJ387747.1 GI:6562532			
KEYWORDS statin.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
Eutheria; Primates; Catarrhini; Homidae; Homo.			
1 (bases 1 to 2512)			
Verheijen,F.W., Verbeek,E., Aula,N., Beereus,C.E.M.T.,			
Harrelsen,M.C., Joosse,M., Peltonen,L., Aula,P., Galjaard,H., Van			
der Spek,P.J.V.D. and Mancini,G.M.S.			
A new gene, encoding an anion transporter, is mutated in statin			
acid storage diseases			
Nat. Genet. 23 (4), 462-465 (1999)			
2 (bases 1 to 2512)			
Verheijen,F.W.			
Direct Submission			

[illegible]

LOCUS 140028 2716 bp DNA PAT 13-MAY-1997
 DEFINITION Sequence 1 from patent US 5618677.
 ACCESSION 140028
 VERSION 140028.1 GI:2083033
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2716)
 AUTHORS Nl,B. and Paul,S.M.
 TITLE Human brain sodium dependent inorganic phosphate cotransporter assay
 JOURNAL Patent: US 5618677-A 1 08-APR-1997;
 FEATURES
 source Location/Qualifiers
 1..2716
 /organism="unknown"
 BASE COUNT 504 a 854 c 738 g 620 t
 ORIGIN

alignment_scores:

Quality: 479.00 Length: 397
 Ratio: 2.056 Gaps: 7
 Percent Similarity: 58.690 Percent Identity: 28.212

alignment_block:

US-09-391-958-1 x 140028 ..

Align seg 1/1 to: 140028 from: 1 to: 2716

```

12 ValProSerLeuCySerAlaArgTyrGlyLeuValLeuHisph 28
   :::::||||| :::::||||| :::::
941 ATCCCTCAGCTGCGGCTGCTATGCTGCTATC..... 979
28 eCyAsnPhrThrIleAlaGlnAsnValIleMetAsnIleThrMet 45
   ||::: ||| ||:::||||| :::::
980 .....TTGCTGAGATCCTGCAGGGGTTGTAGAGGGGTCACATACC 1022
45 AlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSer 61
   |||
1023 CCGCTCTGC..... 1030
62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLys 78
1030 ..... 1030
78 rLeuProAlaLysSerSerIleLeuGlyGlnPheAlaIleTrpGlu 95
   :::::||||| :::::||||| :::::
1031 .....CATGGATCTGAGCA 1046
95 rGTrpGlyProProGlnIuArgSerArgLeuCySerIleAlaLeuSer 111
   :::::||||| ||:::||||| :::::
1047 AATGGGCCACCTTAGAAGAGAGTGGCTGGCGAGCAGACAGCTTTGT 1096
112 GlyMetLeuLeuGlyCySerPheThrAlaIleLeuIleGlyGlyPheLe 128
   ||| ||::: ||::: ||::: ||::: ||:::
1097 GGTTCCTATGCTGGGGGCTGTGCTGCTCTACAGAGCCCGCGCTGT 1146
128 rGluThrLeuGlyTrpProPheValPheTyrIlePheGlyGlyValGly 145
   ::::: ||::: ||::: ||::: ||::: ||:::
1147 GCAGTACTGAGATGAGCTGTTTCTACGCTACGACGACCTTCGGGA 1196
145 yValAlaCySerLeuLeuTrpPheValValIleTyrAspAspProValSer 161
   ||::: ||::: ||::: ||::: ||::: ||:::
1197 TCTTCTGAGACTGTCTGTGCTGCTCTCTACAGAGCCCGCGCTG 1246
162 TyrProTrpIleSerThrSerGluLysGluTyrIleIleSerSerLeu 178
   ::::: ||::: ||::: ||::: ||::: ||:::
1247 CACCCAGCATCTCGAGAGAGAGGACAGTACATC.....GA 1284
178 sGlnGlnValGlySerSerLysGln.....ProLeu..... 188
   ::::: ||::: ||::: ||::: ||::: ||:::
1285 GATCCCATCGAGAGAGCGCAAAACTCATGAAACCCCTCACAGAGTTTA 1334

```

```

189 .....ProIleAlaMetLeuArgSerLeuProIleTrpSerIleCy 203
   ||| ::::: ||::: ||::: ||::: ||::: ||:::
1335 GCACCTCCCTGGCGGCTTTCACGTATGACAGTACATGACATC 1384
204 LeuGlyCySerPheSerHisGlnTrpLeuValSerThrMetValValTrp 220
   ::::: ||| ::::: ||::: ||::: ||::: ||:::
1385 GTGGCAACTTCTCCCGACGTGACCTTACCTGCTGCTCATCTCCA 1434
220 eProThrTyrIleSerSerValTyrHisValAsnIleArgAspAsnGly 237
   ||| ||::: ||::: ||::: ||::: ||::: ||:::
1435 GCCCGACTACTTCAGAGAGGTTCGGCTTCGAGATCAGCAGAGTAGCC 1484
237 euleuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 253
   ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1485 TGGTGTCCGGCTGCCCGACCTGCTATGACATCATCGTCCATCGGC 1534
254 GlyTyrLeuAlaAspPheLeuThrLysLysPheArgLeuIleThr.. 269
   ||| ::::: ||::: ||::: ||::: ||::: ||:::
1535 GCCCAGATCCGGGACTTCTGCGAGCGCGCATCATGTCACACCA 1584
270 .ValArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu 286
   ||| ||::: ||::: ||::: ||::: ||::: ||:::
1585 CGTGGCAGATGATGATGACTGCGGAGGCTTCGGCATGGAACACCGCTC 1634
286 lValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAlaLeu 302
   ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1635 TGTGTGTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1681
303 lLeuThrLeuSerCySGlyLeuSerThrLeuCyGlnSerGlyLeuTrp 319
   ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1682 CTGCTCTGACCGCTGGCTTACACGCTTCCATCTGAGGTTCAACGT 1731
319 eAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetGlyAla 336
   ||| ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1732 GAACCACTGGAGCATAGCCCGCGCTACCCAGCATCTCATAGGAGCAT 1781
336 eArgGlyPheSerSerIleAlaProValIleValProThrValSerGly 352
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1782 CCACAGCGCTGGGACACATGTGCGCATGTGTCGCCCATCATCTGGGG 1831
353 PheLeuLeuSerGlnAspProGlnPheGlyTrpArgAsnValPhePhe 369
   ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1832 GCCATGACTTAAGACACAGACTCGGGAGAGTGCGCATGCTTCTCAT 1881
369 uLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrIlePheGly 386
   ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1882 TGCCTCCCTGTCACATGAGAGTGTCACTTCTACGCGGCTTTCCT 1931
386 lAlaAspValGlnGluTrpAlaLysGluArgLysLeuThr 399
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1932 CTGGAGAGAGACAGCCGCTGGCAGAGCTGAGAGATGAGC 1972

```

seq_name: gb-pat:140029
 seq_documentation_block:
 LOCUS 140029 2716 bp DNA PAT 13-MAY-1997
 DEFINITION Sequence 3 from patent US 5618677.
 ACCESSION 140029
 VERSION 140029.1 GI:2083034
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2716)
 AUTHORS Nl,B. and Paul,S.M.
 TITLE Human brain sodium dependent inorganic phosphate cotransporter assay
 JOURNAL Patent: US 5618677-A 3 08-APR-1997;
 FEATURES
 source Location/Qualifiers
 1..2716
 /organism="unknown"
 BASE COUNT 504 a 854 c 738 g 620 t

OM of: US-09-391-958-1 to: N_Geneseq_36:* out_format : pfs
Date: May 25, 2000 1:59 PM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frmet+ p2n_model -DEFV=xlp
-O=/c9g2_1/USPTO.spool/US09391958/rnatc_24052000_213330_7207/app-query.fasta.1
-DB=N_Geneseq_36 -QFPT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.500 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsunsu2
-TRANS=human40.cdt -LIST=45 -DOCCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTEXT=pfs -NORM=ext -MINLEN=0
-MAXLEN=100000 -USER=US09391958 -NCPU=6 -ICPU=3 -NO_XLPRV -WAIT
-THREADS=1

Search information block:
Query: US-09-391-958-1
Query length: 401
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 62.680000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:V43711	+	2076.00	3760.06	7.8e-202	1643 Human sodium-dependent phosphatase
N_Geneseq_36:V57910	+	2061.00	3731.66	3.0e-200	1195 Human haemochromatosis gene NT
N_Geneseq_36:V7909	+	932.00	1667.68	2.7e-85	2281 Human haemochromatosis gene NT
N_Geneseq_36:T42064	+	479.00	838.63	4.1e-39	2716 Human brain Na+ dependent ionot
N_Geneseq_36:V33503	+	479.00	838.63	4.1e-39	2716 Human sodium-lithium countertr
N_Geneseq_36:V43712	+	433.00	781.08	6.6e-36	2722 Sodium-dependent phosphate cotr
N_Geneseq_36:V57926	+	318.00	493.46	6.9e-20	225033 Hereditary haemochromatosis
N_Geneseq_36:V7903	+	318.00	493.46	6.9e-20	237326 Hereditary haemochromatosis
N_Geneseq_36:T62652	+	160.50	278.13	5.3e-09	1840 DNA encoding protein involved
N_Geneseq_36:T62653	+	150.50	271.63	1.6e-07	3217 DNA encoding protein involved
N_Geneseq_36:V74619	+	150.50	239.49	9.7e-06	2576 Staphylococcus aureus contig S
N_Geneseq_36:V74423	+	138.50	194.87	0.0030	18613 Staphylococcus aureus contig
N_Geneseq_36:V52165	+	131.00	188.04	0.0071	10240 Staphylococcus aureus contig
N_Geneseq_36:V74708	+	124.00	212.08	0.0003	415 Staphylococcus pneumoniae gen
N_Geneseq_36:X07357	+	122.50	195.49	0.0027	5385 Staphylococcus aureus tetracy
N_Geneseq_36:X13329	+	120.50	176.25	0.0323	2288 Staphylococcus aureus genome
N_Geneseq_36:X13024	+	118.50	182.43	0.0146	32768 Staphylococcus aureus genome
N_Geneseq_36:X13060	+	115.50	146.38	1.49	3225 Staphylococcus aureus genome
N_Geneseq_36:X13233	+	113.50	169.36	0.0783	3181 Staphylococcus aureus contig S
N_Geneseq_36:V74611	+	112.50	167.69	0.0969	9271 Staphylococcus aureus tetracy
N_Geneseq_36:V23077	+	112.50	155.41	0.4686	10974 Staphylococcus aureus genome
N_Geneseq_36:V23080	+	112.50	153.47	0.6007	2584 Staphylococcus aureus genome
N_Geneseq_36:N60055	+	112.00	169.17	0.0802	7296 Staphylococcus aureus genome
N_Geneseq_36:T51330	+	112.00	157.25	0.3701	9333 Staphylococcus aureus genome
N_Geneseq_36:V23084	+	111.50	153.51	0.5980	9256 Staphylococcus aureus genome
N_Geneseq_36:V23075	+	110.50	151.78	0.7467	9391 Staphylococcus aureus genome
N_Geneseq_36:V23077	+	110.50	151.78	0.7467	9391 Staphylococcus aureus genome
N_Geneseq_36:T04247	+	109.50	150.12	0.9105	9034 Staphylococcus aureus genome
N_Geneseq_36:T04247	+	109.50	150.12	0.9105	9034 Staphylococcus aureus genome
N_Geneseq_36:T42920	+	109.50	150.12	0.9236	9122 Staphylococcus aureus genome
N_Geneseq_36:T76929	+	109.50	149.78	0.9639	9390 Staphylococcus aureus genome
N_Geneseq_36:T76931	+	109.50	149.78	0.9639	9390 Staphylococcus aureus genome
N_Geneseq_36:V2179	+	109.00	151.34	0.7900	2885 Staphylococcus aureus genome
N_Geneseq_36:T35417	+	107.50	159.46	1.48	9126 Staphylococcus aureus genome
N_Geneseq_36:T45828	+	107.50	146.44	1.48	9126 Staphylococcus aureus genome
N_Geneseq_36:T96074	+	107.50	146.44	1.48	9126 Staphylococcus aureus genome
N_Geneseq_36:V23076	+	107.50	146.13	1.54	9391 Staphylococcus aureus genome
N_Geneseq_36:T76930	+	107.50	146.13	1.54	9391 Staphylococcus aureus genome
N_Geneseq_36:T58840_0	+	107.50	117.87	57.77	110000 Staphylococcus aureus genome
N_Geneseq_36:V23082	+	107.00	146.19	1.53	8630 Staphylococcus aureus genome
N_Geneseq_36:V12391	+	106.50	161.50	0.2146	2102 Staphylococcus aureus genome

N_Geneseq_36:T515656 + 106.50 144.31 1.95 9391 HGV-PNF 2161 polyprotein co
N_Geneseq_36:T94168 + 106.50 144.31 1.95 9391 Hepatitis G virus PNF-2161
N_Geneseq_36:V23083 + 106.50 144.31 1.95 9391 Fragment K3732(21) of a new
N_Geneseq_36:T08812 + 106.50 144.31 1.95 9392 Hepatitis virus open readin

seq_name: N_Geneseq_36:V43711

seq_documentation_block:

ID V43711 standard; DNM: 1643 BP.
AC V43711:
DT 30-NOV-1998 (first entry)
DE Human sodium-dependent phosphate cotransporter coding sequence.
KW Human sodium-dependent phosphate cotransporter; human; NAFTR; cancer; myopathy;
KW cell signalling disorder; phosphate regulation disorder; therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 235..1440
FT /tag= a
PD 27-AUG-1998.
PF 24-FEB-1998; U03745.
PR 24-FEB-1997; US-805118.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Lal P;
DR WPI: 98-467561/40.
DR P-PSDB: W69971.
PT New isolated human sodium-dependent phosphate co:transporter - used
PT to develop products for treating e.g. cancers, osteoporosis,
PT Alzheimer's disease, diabetes, encephalopathy, myopathy,
PT hypocalcaemia or hypoglycaemia.
PS Claim 5: Fig 1: 66pp: English.

CC This sequence encodes the human sodium-dependent phosphate cotransporter
CC (NAFTR) of the invention. NAFTR and agonists of it can be used to treat
CC or prevent disorders associated with decreased phosphate levels,
CC e.g. cancers of the kidney, disorders of decreased phosphate levels
CC including tumoral calcinosis, osteomalacia, osteoporosis, familial
CC hypophosphataemia, rickets, cystineuria, nephrocalcinosis,
CC glomerulonephritis, renal calculus, Alzheimer's disease, diabetes
CC mellitus, hereditary amyloidosis, myopathies including progressive
CC external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic epilepsy,
CC encephalopathy, cardiomyopathy, hypokalaemia, Goodpasture's syndrome, and
CC disorders of cell signalling through cAMP, ATP, NADPH and
CC glucose-6-phosphate. Antagonists or inhibitors of NAFTR may be
CC administered to a subject to treat or prevent disorders associated with
CC increased phosphate levels, e.g. hypocalcaemia, hypocalcaemia, and
CC abnormal phosphate regulation in neurons, gastrointestinal tract and
CC liver. The products can also be used for detection, diagnosis and drug
CC screening.
SQ Sequence 1643 BP; 435 A; 376 C; 359 G; 473 T;

alignment_scores:

Quality: 2076.00 Length: 401
Ratio: 5.177 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x V43711 ..
Align seg 1/1 to: V43711 from: 1 to: 1643

1 MetGlnValAspGluThrLeuIleProArgLysValProSerLeuGlyCys 17
235 ATCCAGATGATGATGACACATGATCCCGAGAAAGTTCCAGATTATGTC 284
17 TAlaArgTyrGlyIleLeuAlaValLeuHisPheCysAsnPhenThrTrp 34
285 TCGTCGCTTATGCAATACCCCTCGCTTACATTTCTGCAATTTTCAACGA 334
34 LeuAlaGlnAsnValIleMetAsnIleThrMetValAlaMetValAsnSer 50
335 TAGACAAATATGATCATCATGACATCCATGATGATGATGATGATGATG 384

51 ThisSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
 385 ACAAGCCCTCAATCCAGACTCAATGATTCCTGAGGCGCTGTTGA 434
 67 PserPhgGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSers 84
 435 CTCATTGGTGCTTAAGTAAAGCCCAAGAGCTTCCTGCAAGTCCT 484
 84 erileuGlyGlyGlnPheAlaIleTrrpGluAqrTrpGlyProGln 100
 485 CAATACCTGGGGGCTCACTTTGCAATTTGGCAAGAGTGGGGCCCTCCACAA 534
 101 GluArgSerArgLeuGlySerLysAlaLeuSerGlyMetLeuGlyGly 117
 535 GAACGAGAGCAGACTGACACATGCTTTATCAGAAATGTTACTGGGATG 584
 117 sPheThrAlaIleLeuIleGlyGlyPheIleSerGlyTrhLeuGlyTrp 134
 585 CTTTACTGCCATCCTCATAGGTGGCTCATAGTAAAGCCCTGGGTGGC 634
 134 robPheValPheTyrIlePheGlyGlyValGlyCysValCysCysLeuLeu 150
 635 CCTTGGCTCTATATCTTGGAGGTGGTGGCTGCTGCTGCTGCTTCTC 684
 151 TrpPheValValIleTyrAspAspProValSerTyrProTrrpIleSerTh 167
 685 TGGTTTGTGTGATTATGATGATGACCCCTTCTATCATCAGATGAAGCAC 724
 167 rSerGlyLysGlyTyrIleIleSerSerLeuLysGlnGlnValGlySers 184
 735 CTCAGAAAAAGATACATCATATCTCTTGAACAAACAGGTGGGTCTT 784
 184 erLysGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTrp 200
 785 CTAGAGAGCTCTCCATCAAAAGTATGCTCAGATCTACACCATTTGG 834
 201 SerIleCysLeuGlyCysPheSerHISGlnTrpLeuValSerThreVal 217
 835 TCCATATGTTTAAAGCTGTTTACCCATCAATGTTTATGTTAGCACAAATGCT 884
 217 lValTyrIleProThrTyrIleSerSerValTyrHisValAsnIleArgA 234
 885 TGTATACATACCAACTTACATCAGCTCTGTATACCATGTTAATCATCAGAG 934
 234 sPAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
 935 ACAATGAGCTTCTATCTGCCCTTCTTATTTGTTGCCGTGATACAGGC 984
 251 MetValGlyGlyTyrIleuAlaAspPheLeuThrLysLysPheArgLe 267
 985 ATGGTGGAGGCTATCTGCAAGATTTCTCTTACCAAAAAGTTTAGACT 1034
 267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSerA 284
 1035 CATCACTGTGAGGAAAAATGCCACAATTTTAGAAGTCCCTCTTTCAG 1084
 284 lAlenIleValSerLeuProTyrIleAsnSerGlyTyrIleThrAlaThr 300
 1085 CACTATATGTCTCTGCTTACCTCAATTCGGGTATATACACCAACT 1134
 301 AlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGlyI 317
 1135 GCCTTGTGAGGCTCTCTTGGGATTTAGACATTTGTGCTGAGTGGGANT 1184
 317 eTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetG 334
 1185 TTATATCATATGCTTAGATATATGCTCAAGGATATCCAGTTTTCATG 1234
 334 lYAlaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
 1235 GAGCATCAAGAGATTTTCAGACATAGCACCTGCTATTGTCACCACTGTC 1284
 351 SerGlyPheLeuLeuSerGlnAspProGlnPheGlyTrpArgAsnValPh 367

seq_name: N_Geneseq_36:V57910
 seq_documentation_block:
 ID V57910 standard; cDNA; 1795 bp.
 AC V57910.
 DT 21-DEC-1998 (first entry)
 DE Human haemochromatosis gene NTP4.
 KW Bovine butyrophilin; BR; human hereditary haemochromatosis; HFE;
 diagnosis; iron metabolism; NPT3; NPT4; Kofret; BTF1; BTF2; BTF3;
 BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 type 1 sodium transport gene; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 377..1582
 FT /tag= a
 FT /product= "NTP4"
 PN W09814466-A1.
 PD 09-APR-1998.
 BP 30-SEP-1997; U17658.
 PR 07-MAY-1997; US-852495.
 PA 01-OCT-1996; US-724394.
 PA (PROG-) PROAGENTOR INC.
 PI Feder JN, Krommal GS, Laufer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 DR P-PSDB: W78920.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Claim 58: Fig 7. 209pp. English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC presence of the HFE gene mutation in the genome of the individual.
 CC The HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophilin (BTF), and can be used in the production of agonists
 CC and antagonists of BTF function. Also described are: (1) a Kofret gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence encodes
 CC NTP4.
 SQ Sequence 1795 BP; 473 A; 436 C; 402 G; 484 T;
 alignment_scores:
 quality: 2061.00 length: 401
 Ratio: 5.165 gaps: 0
 Percent Similarity: 99.501 Percent Identity: 99.252
 alignment_block:
 US-09-391-958-1 x V57910 ..

Align seg 1/1 to: V57910 from: 1 to: 1795

```

1 MetGlnValAspGluThrLeuIleProArgLysValProSerLeuGlySse 17
  |||||||
377 ATGCAAGTGTAGACACTGATGCCAGGAAGTCCAGATTATGTTTC 426
17 AlaAaGlyGlyIleAlaLeuValLeuHisPheCysAsnPheThrThr 34
  |||||||
427 TGTCTGCTATGGAATAGCCCTCCTTACTTCTGCAATTTACAAACA 476
34 LeuAlaGlnAsnValIleMetAsnIleThrMetValAlaMetValAsnSer 50
  |||||||
477 TACACAAAATGTCATCATGACATCACCATGTTAGCCATGATGCAACAGC 526
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
  |||||||
527 ACAAGCCCTCAATCCAGCTCAATGATTCCTCTGAGTCCCTGCTGTGA 576
67 PserPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSerS 84
  |||||||
577 CTCATTGTGGGCTAAGTAAGCCCCAAAGAGTCTTCTGCAAAAGCTC 626
84 eRIleLeuGlyGlyGlnPheAlaIleTrpGluArgTrpGlyProProGln 100
  |||||||
627 CAATACTTGGGGGTCAAGTTTGCATTTGGGAAAGTGGGCCCTCCACAA 676
101 GluArgSerArgLeuGlySerLysIleAlaLeuSerGlyMetLeuLeuGly 117
  |||||||
677 GAACGAGAGAGACTTCAGCATTCCTTTATCAGGAATGTTACTGGGATG 726
117 sPheThrAlaIleLeuIleGlyLysPheIleSerGluThrLeuGlyTrp 134
  |||||||
727 CTTTACTGCCATCTCATAGTGGCTTCAATAGTGAACCCCTGGGGGTC 776
134 roPheValPheTrpIlePheGlyIleValGlyCysValCysCysLeuLeu 150
  |||||||
777 CCTTGTGCTCTATATCTTTGGAGGTGGTGGCTGCTGCTGCTGCTCT 826
151 TrpPheValValIleTrpAspAspProValSerTrpProTrpIleSerTh 167
  |||||||
827 TGGTGTGCTGATTTATGATGATGCCCTTTCTCATCAATGATGAACAC 876
167 rSerGluLysGluTrpIleIleSerSerLeuLysGlnGlnValGlySerS 184
  |||||||
877 CTGAGAAAAGATACATCATCTCTCTGAAACAACAGCTGGGGCTT 926
184 eRIYsGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTrp 200
  |||||||
927 CTAGAGAGGCTCTTCCATCAAGCTATGCTCAGATCTCTACCCATTGG 976
201 SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMetVa 217
  |||||||
977 TCCATATGTTTAAAGCTGTTTCAACCAATGATGTTAGTGTAGCAATGCT 1026
217 ValValTrpIleProThrTrpIleSerSerValTrpHisValAsnIleArg 234
  |||||||
1027 TGTATACATACCAACTTACTACGCTCTGTGTACCAATGTTAAACATCAG 1076
234 sPasnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
  |||||||
1077 ACAATGGACTCTATCTGCCCTTCTTTATTTTGTGCTGGGTCATAGGC 1126
251 MetValGlyGlyTrpLeuAlaAspPheLeuLeuThrLysLysPheArgle 267
  |||||||
1127 ATGGTGGGAGGCTATCTGCGAGATTTCTTCAACCAAAAGTTTACGT 1176
267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSera 284
  |||||||
1177 CATCACTGTGAGGAAATTCACCAATTTTAAAGAGCTCCCTCTTTCAG 1226
284 IaleuIleValSerLeuProTrpLeuAsnSerGlyTrpIleThrAlaThr 300
  |||||||

```

```

1227 CACTCATTTGTCTGTGCTTACCTCAATTCGCGCTATATCAGACNACT 1276
301 AlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuGlnSerGlyI 317
  |||||||
1277 GCGTTGTGAGCTCTCTTCCGATTAAGCAATGTGTGATGATGAGGAT 1326
317 eTrpIleAsnValLeuAspIleAlaProArgTrpSerSerPheLeuMetG 334
  |||||||
1327 TTATATCAATGCTTATGATATTTGCTCCAAAGTATTCAGATTCTCTAC 1376
334 ValAserArgGlyPheSerSerIleAlaProValIleValProThrVal 350
  |||||||
1377 GAGCATCAAGAGATTTCAGCATATAGCACCCTGATTCATACCACTGTC 1426
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValPh 367
  |||||||
1427 AGCGGATTTCTTCTTACTAGGACCCCTGATTTGGGGAGGAATGCTT 1476
367 ePheLeuLeuPheAlaValAsnLeuGlyLeuLeuPheTrpLeuIleP 384
  |||||||
1477 CTCCTTGTCTTGTGCGCTTAACCTGTAGGACTACTCTTCTTACCTCAT 1526
384 heGlyGluAlaAspValGlnGluTrpAlaLysGluArgLysLeuThrArg 400
  |||||||
1527 TTGGAGAGAGAGATGTCCAGAAATGGGCTAAAGAGAAACATCACTCT 1576
401 Leu 401
  |||
1577 TTA 1579

seq_name: N_Geneseq_36:V57909
seq_documentation_block:
ID V57909 standard; cDNA; 2281 BP.
AC V57909;
DR 21-DEC-1998 (first entry)
DE Human haemochromatosis gene NRP3.
KW Bovine butyrophilin; Bf; human hereditary haemochromatosis; HFE;
  diagnosis; Iron metabolism; NRP3; NPT4; Roket; BTF2; BTF3;
  BTF4; BTF5; milk protein; Lupus; Sjogren's syndrome; hypophosphatemia;
  type 1 sodium transport gene; ss.
OS Homo sapiens.
FH Key location/Qualifiers
FT CDS 419..1729
FT /tag= a
FT /product= "NTP3"

WO9814466-A1.
09-APR-1998.
30-SEP-1997: U17658.
07-MAY-1997: US-852495.
01-OCT-1996: US-724394.
(PROG-) PROGENITOR INC.
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WD,
  Tsuchihashi Z, Wolff RK;
DR WPI: 98-240014/21.
DR P-PSDB: W78919.
PT Hereditary haemochromatosis gene products - used to develop products
  for the diagnosis and treatment of hereditary disorders in iron
  metabolism
PT Claim 52: Fig 7: 209pp; English.
PS The present invention describes hereditary haemochromatosis gene
  products from the human haemochromatosis gene. Also described is a
  method to determine the presence or absence of the common hereditary
  haemochromatosis (HFE) gene mutation in an individual comprising:
  (a) providing DNA or RNA from the individual; and (b) assessing the
  DNA or RNA for the presence or absence of a haplotype or genotype where
  the presence or absence of the haplotype genotype indicates the likely
  presence of the HFE gene mutation in the genome of the individual. The
  HFE gene sequences from the present invention can be used to develop
  products for use in the diagnosis and treatment of HFE. The present
  invention also describes BTF genes, which are homologues of the milk
  protein butyrophilin (Bf), and can be used in the production of agonists
  and antagonists of Bf function. Also described are: (1) a Roket gene

```

CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence encodes
 CC NPT3.

Sequence 2281 BP; 624 A; 506 C; 490 G; 661 T;

alignment_scores:

Quality: 932.00 Length: 471
 Ratio: 3.305 Gaps: 4
 Percent Similarity: 59.873 Percent Identity: 41.826

alignment_block:

US-09-391-958-1 x V57909 ..

Align seg 1/1 to: V57909 from: 1 to: 2281

```

10 ArgLysValProSerLeuCySerAlaArgTyrGlyIleAlaLeuVal 26
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 AGGAAAGGTCAGATTTCTGTTCATTAAGCTATGGCTGCTTTATCAT 489

26 uHSPheCyAsnPhetThrIleAlaGlnAsnValIleMetAsnIle 43
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 GCACCTCTCAACTTCACATGATTAAGCAGCGTGTGAGCTGACGATG 539

43 hMetValAlaMetValAsnSerThrSerProGlnSerGlnLeuAsn 59
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 CGATCATGCCCATGTGTGAACACACACACAGAGAGAGCTATCTATGCC 589

60 SerSerGluValLeuProValAspSerPheGlyGlyLeuSerLys 76
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 TCCACTGAGGGGCGCTGTTCAGATGCTTCATTAACCTCCAGATATCAT 639

76 ouySerLeuProAlaLysSerSerLe..... 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 CAAGGAATTTGATACAAAGGCTCTGTGTATCAATGAGGCCAGAACTC 689

85 ..... 85
690 AGGATATCATCTTAGCTTCATCACTATGAGATATACTGACTGTGATC 739

86 ..... 87
740 CCAATGGATATTAGACAGGATATTGGAGCAAAAAAATGCTGGTGC 789

87 ..... 87
790 TGGTTGCTGATCTTCCCTTCTCACCCCTCTTTACACACGCTGCTG 839

87 ..... 87
840 ACTTGGAGATGATTTTGGTCATCATGTTCCGACAGTCAGGGCATGCGC 889

88 ..... 98
890 CAGGAAATGGCATGACAGGTCACTTACTATTGGGCAAAAGTGGGCTCC 939

98 oProGlnIuArgSerArgLeuCySerIleAlaLeuSerGlyMetLeu 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
940 TCCACTTGAGCAAGACAGCTCACACCATTCAGAGATCAGGTCAGCAT 989

115 euGlyCyAspPhetThrAlaIleLeuIleGlyGlyPheIleSerGln 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
990 TTGATCTCTTCATCATCTCTGTGTGGGGGACATTAATCTCAGAGCCCTG 1039

132 GlyTrpProPheValPheTrIlePheGlyGlyValGlyCyValCy 148
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1040 AGCTGGCCTTTATCTTCTACATCTTGTGACACTGGCTGTGCTGCTG 1089

148 sLeuLeuTrpPheValValIleTyrAspAspProValSerTyrPro 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1090 TCTCTATGTTGTCAGATGATTATGATGACCCATCATCACCCGTCGA 1139

```

```

165 LeSerThrSerGluGlyTrpIleIleSerSerLeuGlnGlnVal 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1140 TAAGTGTAGGAAAGAGAGACATCTGCTCCATGCTGCTCAACAGCCC 1189

182 GlySerSerLysGlnProLeuProIleValIleMetLeuArgSer 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1190 AGTTCTCTGAGACGAGCTGCCCATTAAGGCGATGTCATGCTCAACC 1239

198 oIleTrpSerIleCyLeuGlyCyAspSerHisGlnTrpLeuValSer 215
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1240 ACTTGTGGCCATTTCTCTGCTGCTTTTTCAGCCATTTCTGTATGACCA 1289

215 hMetValValIleTrpThrIleSerSerValTyrHisValAsn 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1290 TCATCTTAACATACCTACCAACGATATATCATGCTGCTCATTTAAC 1339

232 IleArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAla 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1340 ATCAGAGATAGTGAAGTCTGTCTCCCTGCTTTATTTGCTGCTGCAAG 1389

248 IleGlyMetValGlyGlyTyrLeuAlaAspPheLeuThrLys 264
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1390 CTGTACAATTTTAGAGAGCTGACCTGACGATTTCTTTGTCCAGGAATC 1439

265 .. PheArgLeuIleThrValArgLysIleAlaThrIleLeuGly 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1440 TTTCTAGATGTATCATCTGTCCAAAGCTTTTTCATCTCTTGA..... 1482

281 ProSerSerAlaLeuIleValSerLeuProTyrLeuAsnSerGly 297
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 ..... 1482

297 eThrAlaThrAlaLeuLeuThrLeuSerCyGlyLeuSerThrLeu 314
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 ..... 1482

314 InsSerGlyIleTyrIleAsnValLeuAspIleAlaProArgTyr 330
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1483 ..... 1491

331 PheLeuMetGlyAlaSerArgLysPheSerSerIleAlaProValIle 347
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1492 TTCTCATGGGAATCTCAAGGGATTTGGCTCATCGAGAAATCATCTC 1541

347 lProThrValSerGlyPheLeuLeuSerGlnAspProGluPheGly 364
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1542 TTCACACTGCCACTGATTCCTCATCAGTCAGATTTTGAAGTGTGTTGA 1591

364 rGAsnValPhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuPhe 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1592 GGATGTCTTTTCTCTGTCTGCTGCAACATGATTTGGCTGCTGCTTT 1641

381 TyrLeuIlePheGlyGlyAlaAspValGlnIuTrpAlaLysGly 397
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1642 TACCTACGTTTGGACAAGAGAACTTCAAGACAGGSCCAAAGAGAGAC 1691

397 sLeuThrArgLeu 401
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1692 CTTTACCGGCTC 1704

seq_name: N_Geneseq_36:T42064
seq_documentation_block:
ID T42064 standard; cDNA: 2716 BP.
AC T42064;
DT 30-JAN-1997 (first entry)
DE Human brain Na+ dependent inorganic phosphate cotransporter cDNA.
KW Brain sodium-dependent inorganic phosphate cotransporter; hbnpt;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 461..2143

```


PD 27-AUG-1998.
 PF 24-FEB-1998: U03745.
 PR 24-FEB-1997: US-805118.
 PA (INCYTE) INCYTE PHARM INC.
 PI Bandman O, Lai P.
 DR WPI: 98-467561/40.
 PT New isolated human sodium-dependent phosphate co:transporter - used
 PT to develop products for treating e.g. cancers, osteoporosis,
 PT Alzheimer's disease, diabetes, encephalopathy, myopathy,
 PT hypocalciuria or hypoglycaemia.
 PS Example V; Page 47; 66pp; English.
 CC This sequence represents a fragment of the DNA encoding the human
 CC sodium-dependent phosphate cotransporter (NAPTR) of the invention. NAPTR
 CC and agonists of it can be used to treat or prevent disorders associated
 CC with decreased phosphate levels, e.g. cancers of the kidney, disorders of
 CC decreased phosphate levels including tumoral calcinosis, osteomalacia,
 CC osteoporosis, familial hypophosphataemia, rickets, cystinuria,
 CC nephrocalcinosis, glomerulonephritis, renal calculus, Alzheimer's
 CC disease, diabetes mellitus, hereditary amyloidosis, myopathies including
 CC progressive external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic
 CC epilepsy, encephalopathy, cardiomyopathy, hypokalaemia, Goodpastures
 CC syndrome, and disorders of cell signalling through cAMP, ATP, NADPH and
 CC glucose-6-phosphate. Antagonists or inhibitors of NAPTR may be
 CC administered to a subject to treat or prevent disorders associated with
 CC increased phosphate levels, e.g. hypocalciuria, hypocalcaemia, and
 CC abnormal phosphate regulation in neurons, gastrointestinal tract and
 CC liver. The products can also be used for detection, diagnosis and drug
 CC screening.
 SQ Sequence 272 BP; 67 A; 56 C; 58 G; 91 T;

alignment_scores:
 Quality: 433.00 Length: 85
 Ratio: 5.094 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-391-958-1 x V43712 ..

Align seg 1/1 to: V43712 from: 1 to: 272

```

317 11etrrlleasvallyleuaspllealeaProArGtyrserSerPheleume 333
|||||
1 ATTTATATCATGCTTATGATATTTGCTCCAAGATATTCAGTTTTCAT 50
333 tgllyalaserarglypheserSerilleaProvalillevalProthrv 350
|||||
51 GGGAGCATCAAGAGATTTCGAGCATAGACCTGTCATTGTACCACCTG 100
350 alserglyPheleuleuSerGlnAspProGluPheglyTPAAsnVal 366
|||||
101 TCAGTGGATTTCCTTCTAGACACCTCAGTTGGTGAGAGAAAGTC 150
367 PhePheleuleuPhealaValasnleuleGlyleuPheyleu 383
|||||
151 TTCTCTCTGCTGTTGCCGTTAACCTTTAGGACTACTCTTCACTCAT 200
383 epheglyGluAlaAspValGlnGluTPAlaLySGluArgLysLeuThra 400
|||||
201 ATTTGAGAGACAGATGTCCAGAAATGGCTAAAGAGAAACTCACTC 250
400 rgleu 401
|||||
251 GTTTA 255

```

seq_name: N_Geneseq_36: V57926

seq_documentation_block:

ID V57926 standard; DNA; 235033 BP.

AC V57926;

DR 23-DEC-1998 (first entry)

DE Hereditary haemochromatosis subregion from an unaffected individual.

KM Bovine butyrophillin; BF; human hereditary haemochromatosis; HFE;

KM diagnosis; Iron metabolism; NPT3; NPT4; RORet; BTF1; BTF2; BTF3;
 KM BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 KW type 1 sodium transport gene; ss.
 OS Homo sapiens.
 PN WO9814466-A1.
 PD 09-APR-1998.
 PF 30-SEP-1997: U17658.
 PR 07-MAY-1997: US-852495.
 PR 01-OCT-1996: US-724394.
 PA (PROG-) PROGENITOR INC.
 PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Example 2; Fig 8; 209pp; English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. The present sequence
 CC represents a hereditary haemochromatosis (Hh). Also described is a
 CC unaffected by hereditary haemochromatosis (Hh). Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the haplotype genotype indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophillin (BTF), and can be used in the production of agonists
 CC and antagonists of BTF function. Also described are: (1) a RORet gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia.
 SQ Sequence 235033 BP; 68786 A; 48466 C; 49441 G; 68340 T;

alignment_scores:
 Quality: 318.00 Length: 543
 Ratio: 1.975 Gaps: 9
 Percent Similarity: 29.650 Percent Identity: 20.994

alignment_block:
 US-09-391-958-1 x V57926 ..

Align seg 1/1 to: V57926 from: 1 to: 235033

```

137 Pherylllepe.....glyGlyValGlycysValcysCysleuleutr 151
|||||
217123 TTTATATCCCTCTCCAGGATGACCTGGCTGCTCTCTCTCTATG 217172
151 pphvalvalletrraspaspProvalserTyProtrPleserhrs 168
|||||
217173 GTTACAGTGTATTTATGATGACCCCATGACATCACCGGCTAAAGTGT 217222
168 erglulysGluTyrilleleserSerleuysGlnGlnValGly..... 182
|||||
217223 GGGAAAGAGACACATCTCTGCTCAGCTGCTCAACAGGTACATGCACA 217272
183 .....Se 183
217273 CCTTGACTCTGCGCCCATGACAGAGTCTCTAGGGCAGGCTGTGATCTC 217322
183 rserlysgln..Pro..... 187
217323 CTCTGAGGAGCACCATTCTTGCGCTCTCTAATATCATGCTGATTAGATCT 217372
188 .....leuproillelysalamelt 194
217373 TTCTTTTACGCCAGTTCTCTCTGACGAGCTGTCCCATTAAGGCGATGG 217422

```

```

218333 ATCTCTAATACACTACTGTGTAAACTCTCTCTCCAGTGCAGTGGAGTCTGTG 218372
239 SerAlaLeuProPheIleValAlaIlePValIleGlyMetValGlyGlyTY 255
218373 TCTCTCCCTGCTCTTTTATGTGCTGTGCAGACCTGTACAAATTTTAGAGAGGCA 218422
255 rLeuAlaAspPheLeuLeuLeuThrTrpLys...PheArgLeuIleThrValA 271
218423 GCTGGCAGATTCCTTTGTGCCAGGAATCTTCTCAGATTGATCATCTGTGC 218472
271 rGlyIleAlaIleThrIleLeuGlySer.Leu..... 280
218473 GAAAGCTCTTTTCATCTCTTGGTAAGGATAACGCTGTGGCCCATTTAAC 218522
280 ..... 280
218523 CAATCCCTTTTTCGACATGTCTCAGAGGGTCCCTGCAGCAGATGCT 218572
281 .....ProSer.SerAlaLeuIleValSerLeuProTY 291
218573 CATTTGCCAGAGGCTCTCTCTCCATCAMPATNGTGTGGCCCTGCCCCCTT 218622
291 rLeuAsnSerGlyTYrIleThrAlaIleThrAlaLeuLeuThrLeuSerCysC 308
218623 TGTGGCTCCAGTTACGTGATACCATTAATTTCTGATCTACTTATTCCTG 218672
308 rLeuSerThrLeuLeuCysGlnSerGlyIleTYrIleAsnValLeuAspPle 324
218673 GGACCAAGTACCACTATGTGACTATGAGGTTTATTCATCAACACCTTAGATATC 218722
325 AlaProArgTYrSerSerPheLeu 332
218723 GCCCCAGGTAGAGCTCTACCTG 218746
seq_name: N_Geneseq_36:v57903
seq_documentation_block:
AC V57903 standard; DNA; 237326 BP.
AD V57903;
DT 21-DEC-1998 (first entry)
DE Hereditary haemochromatosis subregion from an HH affected individual.
KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW diagnosis; Iron metabolism; NP3; NP4; Kolet; BR1; BR2; BR3;
KW BR4; BR5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene; ss.
OS Homo sapiens.
PN MO9814466-A1.
PD 09-APR-1998
PE 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PA 01-OCT-1996; US-724394.
PA (PROG-) PROGENERATOR INC.
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
PI Wredt; 98-240014/21.
DE Hereditary haemochromatosis gene products - used to develop products
PI for the diagnosis and treatment of hereditary disorders in iron
PI metabolism
PS Claim 1; Fig 9; 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. The present sequence
CC represents a hereditary haemochromatosis subregion from an hereditary
CC haemochromatosis (HH) affected individual. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes pmf genes which are homologous to the HFE gene

```

us-09-391-958-1.rng

220040 CATCATATTTCTTAACCACTATGCTTACTACCAACACAGCTATTCCAA 220085
234 234
220090 AGCTTCTTTTAGAAATAATATGCTGCGCAGCAGTGGCTCATGCCCT 220135
234 234

234 234

220130 ATGCAAGACCAAGCTGCACCAATATGCTTACATTAATATCATCTACTATAAA 220239

234	234
-----------	-----

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

220290 CAGGAGGCTGAGACAGAGAATCGCTTGAACCCAGAGGTGAGGTTGCA 220339

220340 TTGAGCCAGATCATGCCACTGCACCTCCAGCTGGGCGACAGAGTAAGAC 220389

220390 TCCGTTTCAAAAACCAAGAAATTATATTGCTTTTATCTGGA 220439

220440 GCCCAGATGATGACCTTCTGGCCCTTATATCTGAGACAGTGTCTTTT 220489

220490 AGTGTGAAAAAGATGCTAATTTTCCCCCAACAACCCACAGTATCATGG 220539

220540 GGGTAAGTTAATGCGCTGCTGTGTAACGACAATTTTGGTGCTAACGT 220589

233AsnGlyLeuLeu 238
220590 ATCTTATACTACTCTGTATAACTTCTCTCTCAGAGTGGAGTTCG 220639

239 SeiaIaleuProphellievaIAIatrpvalleiglymetValgylty 255
|||::|||||||::||| ::::|
220640 TCCTCCCTGCCTTTATTCGTGCAAGCTGTACAATTTTAGAGGTCA 220689

255 rlu61aasphneleuennrLysLys...pheargleuennrVala 2/1
 |||||:::|||||
 220690 GCTGGCAGATTCCITTTTGTCCAGGAATCTTTCAGATTGATCATCTGTGC 220735

220740 GAAAGCTCTTTCATCTCTTGGTAAGGATTAAGCGTGCGGCCCATTTAAC 220789

280 280

220790 CAATCCCTTTCTGCACATGGTCTCAGAGGTTCCCTGCACAGATGTCCCT 220839

```

281 .....Proser.SerialLeuIleValSerLeuProty 291
220840 CATTGCCCAGGGCTCTCTTCATCAATATGTGCTGTGGCCCTGCCCCTT 220889

```

291 rluasanserglytyrilethrAlathralaleuleuthrlseiscysg 308
::: |||::|||::: :: | |||| |
220890 TGIGGCCCTCAGTTACGTGATAACCATTATTTTGCATACATTATTCCTG 220933

308 IyeuSerThrIeucysInserglyIleeyrIIeaSnValIeuaspIle 324
|||:::||||:::||||::: ||||:::||||||
220940 GCACCAGTAACCTATGTGACTCAGGGTTTATCATCAACACCCTTAGATATC 220989


```

372 aValAsnLeuGlyLeuLeuPheTyrlLeuIle...PheGlyGlyAla 387
||||| ||||| ::||| ||||| ::||| |||||
1491 CGTGCCCTCATACCTGTGCCTTCCATTATCATCCGCTATGGAGAGGGG 1539
seq_name: N_Geneseq_36:762653

seq_documentation_block:
ID 762653 standard; DNA; 3217 BP.
AC 762653;
DT 15-MAY-1997 (first entry)
DE DNA encoding protein involved in cephalosporin C biosynthesis.
KM cephalosporin C; biosynthesis; fermentation; ds.
OS Acromonium chrysoeum.
FH key location/Qualifiers
FT exon 935..1101
FT /tag- a
FT /note- "ATG start codon commences at nucleotide 935,
FT actual exon start not indicated in specification"
FT Intron 1102..1157
FT /tag- b
FT exon 1158..1614
FT /tag- c
FT Intron 1615..1675
FT /tag- d
FT exon 1676..2245
FT /tag- e
FT Intron 2246..2302
FT /tag- f
FT exon 2303..2602
FT /tag- g
FT /note- "actual exon end not indicated in specification,
FT nucleotide indicated here is end of coding
FT sequence, TGA stop codon found at 2603-2605"
PN 109009966-A.
PD 14-JAN-1997.
PE 03-JUL-1995; 167461.
PR 03-JUL-1995; JP-167461.
PA (ASAH ) ASAH KASEI KOGYO KK.
PFI 97-126424/12.
DR P-PSDB; M14439.
PT Acromonium chrysoeum gene involved in biosynthesis of
PT cephalosporin C - used to improve fermentation ability of A.
PT chrysoeum
PS Claim 4: Page 11-14; 21pp: Japanese.
CC This DNA, derived from Acromonium chrysoeum, contains a coding
CC sequence for a protein involved in biosynthesis of cephalosporin C.
CC The gene involved in biosynthesis of cephalosporin C or its cDNA
CC can be used to improve the fermentation ability of Acromonium
CC chrysoeum.
SO Sequence 3217 BP; 683 A; 934 C; 769 G; 798 T; 1 U;

alignment_scores:
Quality: 169.50 Length: 430
Ratio: 0.807 Gaps: 24
Percent Similarity: 48.837 Percent Identity: 23.953

alignment_block:
US-09-391-958-1 x 762653 ..
Align seg 1/1 to: 762653 from: 1 to: 3217

```

```

97 lYPro.....ProGlnIuArgSerArgLeuS 106
||| ||||| ||||| ||||| ||||| |||||
1502 GTGCCACCGTCGCCGATCTTTTATCCAGAGAGCGGCTGGCCATG 1551
107 SerIleAlaLeuSerGlyMetLeuLeuGlyCysPheThrAlaIleLeu1 123
::||| ||||| ||||| ||||| ||||| |||||
1552 GCCATTGGAGCTTTGGCCCGTTGATGGGTCTCGTGTGGTCCGCGT 1601
123 eGlyGlyPheIleSerGlyThrLeuGlyTrpPheValPheTyrlLe 139
::||| ||||| ||||| ||||| ||||| |||||
1602 GGGAGGATATCTTGGCCAGGCCAGGTTGGCGGCGTGTGGTGGTGG 1651
140 ....PheGlyGlyValGlyCysValCysLeuLeuTrpPheValAl 154
1652 TCGCCATCGGGGTACGTATGTGTTCATTGTGAGATCGTTCTCCAG 1701
155 IleTyraAsp.....ProGlnIuArgSerArgLeuS 158
||| ||||| ||||| ||||| ||||| |||||
1702 GTTTCAGCTGACCGCATGTCTAGGGGGCTTCATACAGCATGTCTTC 1751
159 .....Pro.ValSerTyrrProTrpIleSerThrSerGlyGlyTr 172
||| ||||| ||||| ||||| ||||| |||||
1752 TCATAGCCCGGTGAGACCTATCCCGGTTCTCCACAGCGCAG..... 1795
173 IleIleSerSerLeuGlyGlnGlnValGlySerSerGlnProLeu.. 188
::||| ||||| ||||| ||||| ||||| |||||
1796 ..GTGAACCGCTACGACAGAAACCGGCAT.....CCCTCTCT 1833
188 .....ProGlnIuArgSerArgLeuS 188
1834 TACCTACACACTAGCGGACCTCGATCGAGCGCTCGATTCCCGGT 1883
189 .....ProIleLysAlaMetLeuArgSerLeuProIle..... 199
||||| ||||| ||||| ||||| ||||| |||||
1884 CTGTCGCCGCGCCCTCGTCTCTCTCCGCTCA...CCCATCGCTTC 1930
200 ...TrpSerIleCysLeuGly...CysPheSerHisGlnTrpLeuAl 214
::||| ||||| ||||| ||||| ||||| |||||
1931 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1975
214 TrpMetValAlaTyrlleProThrTrpIleSerSerValTrpHis 231
::||| ||||| ||||| ||||| ||||| |||||
1976 ...CTCTTGTGCACTATCCCAAGTGTCTTGGCAGATCTACGACT 2021
231 snIleArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAl 247
::||| ||||| ||||| ||||| ||||| |||||
2022 CGCTGTGTCATAATAGC...CTCTCTATCTGGGATATGGCGGCGCT 2068
248 ValIleGly...MetValGlyGlyTyrlleAlaAspPheLeuThr 263
::||| ||||| ||||| ||||| ||||| |||||
2069 CTGCTGGGGAATGCCATATTCGGCCAGCTCTCTCTCTCTCTCT 2118
263 sLys.....PheArgLeuIleThrVal 271
||| ||||| ||||| ||||| ||||| |||||
2119 AAAGTCTGGCGGATGAGTTGAACCGGAGTTCTCTCTCTGATG. 2167
271 rGlyIleAlaThrIleLeuGlySerLeuProSerSerAlaLeuIle 287
::||| ||||| ||||| ||||| ||||| |||||
2168 .....ATCCC...GGGCTTTTGTGATC 2188
288 SerLeuProTyrlleuAsnSerGlyTyrlleThr..... 298
::||| ||||| ||||| ||||| ||||| |||||
2189 CCCATGTGTTCATCTACGGGCGGACCTACTCAAGCTCCACTG 2238
299 .....AlaThrAlaLeuLeuThrLeuSerGlyLeuS 310
2239 GATGATGCCCATCTGTGAGGTGCTCTCTGCGGATTTGACCTCAAC 2288
310 eT..... 310
2289 CACGCTGAAGTCTTCTCTGCACTGGGATTTGAAACACCGCTCTAC 2338

```


Key Location/Qualifiers
misc_feature 901..960
/tag= a
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 2701..2760
/tag= b
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 4501..4560
/tag= c
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 6301..6360
/tag= d
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 8101..8160
/tag= e
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 9901..9960
/tag= f
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 11701..11760
/tag= g
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 13501..13560
/tag= h
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 15301..15360
/tag= i
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
misc_feature 17101..17160
/tag= j
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

EP-786519-A2.
30-JUL-1997.
07-JAN-1997; 100117.
05-JAN-1996; US-009861.
(HUMAN-) HUMAN GENOME SCI INC.
Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
Rosen CA:
MPI: 97-374922/35.
Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -
stored on computer readable medium and used in the production of
anti-S.aureus vaccines
Claim 1; Page 641-651; 3271pp; English.

CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
Sequence 18613 BP; 5990 A; 2753 C; 3477 G; 5782 T;
alignment_scores: Quality: 138.50 Length: 392
Ratio: 0.774 Gaps: 19
Percent Similarity: 45.663 Percent Identity: 20.663
alignment_block:
US-09-391-958-1 x V74423 ..
Align seg 1/1 to: V74423 from: 1 to: 18613
84 SerIleuGlyGlyGlnPheAlaIleTrrp..... 93
||||:|||||.....
9146 TCTGTAAATGGAATTAATTAATGTAATGCGGACTTAACGGCGTGTCCA 9195
94GluArGTrpGlyProp 99
9196 ATCACTTGCTGGACCTGCAAGTATTCAACGATTTCAGAGTGGGCCCA 9245
99 roGInGluArGSerArGleuCySerIleAlaLeuSerGlyMetLeuLeu 115
9246 GAACGAACGT.....GCCGATACCTTA 9268
116 GlyCysPhe...ThrAlaIleuIleGlyPheIleSerGluThrLe 131
9269 GGAATTCGTGAATACATCAATTAATTCGTCGTGCATGACGATGTCGT 9318
131 uGly...TriProPheValPheTyrIlePheGlyGlyVal..... 143
9319 TGCACTTGGGGTGTATATGTAATTCCTCCATGGAATATTTATAGGAGT 9368
144GlyCysValCysCysLeuLeuTrpPheValIleTyr 156
9369 TCATTTCCATCGCGTGAATGATTAATTTGATGCAACATTAATTT 9418
157AspAspProValSerTyrProTrpIleSerTrnSerGlu.. 169
9419 ATCGCAAAAGATGATCGGAGCAAGAAATAGCATTCGTCGAGAAT 9468
170LysGluTyrIleIleSerSerLeuLysGln 180
9469 TTGGCAAGACCGCGATTAAGAAATATGATTCCTCA..... 9508
180 InvalGlySerSerLysGlnProLeuProIleLysAlaMetLeuArGser 196
9509GATATGACGAATAGGAGATCTTAAATAATATATCCTGGAAAT 9553
197 LeuProIleTrpSerIleCysLeuGlyCysPheSerHISGlnTrpLeuVa 213
9554 CCGTTATATGATTCATGAT.....GT 9576
213 IserThreValValTyrIle.....P 221
||||:|||||

[illegible]

polynucleotide sequences - useful in diagnostic kits and assays, and pharmaceutical compositions and vaccines for Streptococcus pneumoniae

Claim 1: Page 326-332: 1409pp: English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (VS2134 to VS2524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (VS2134 to VS2524) are genomic fragments from Streptococcus pneumoniae. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridize to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the S. pneumoniae genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the S. pneumoniae genome of commercial importance, or expression modulating fragments of the S. pneumoniae genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for S. pneumoniae.

Sequence 10240 BP: 2865 A: 1914 C: 2390 G: 3068 T:

[illegible]

alignment_block:
US-09-391-958-1 x V52165

Align seg 1/1 to: V52165 from: 1 to: 10240

[illegible]

```
9530 .GCCAAGAAAAGGCTATTCAACAAGAAATATTACCGCTTAA 9577
199 ILeIrrSerIleCysLeuGlyCysPheSerHisIleTrpLeuValSerTh 215
9578 TATCCCTATCTTTTGGCAATCTCTT.....TTAACCAAGTT 9615
215 mctValValTrIleProThrTrpIleSerSerValTyHisValAsn 232
9616 TGTCATCCAAATTTTCAGCTCAATCGATGGCCCTAATTTGGCTCTTATG 9665
232 leaArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpVal 248
9666 TAGCCGACTTAGGACACAGACAGAAATCTCTTTTGTCTCGTTGATTT 9715
249 ILeGlyMetValGlyGlyTrpLeuAlaAspPheLeuLeuThrIlySlySph 265
9716 GTGTCGCGTATGGCC..... 9730
265 eaTrpLeuIleThrValArgIlySileAlaThrIleLeuGlySerLeuProS 282
9731 .....TTTCCAGCATGATGAGTGCAGAGTCATGCGCAAGCTAGCTG 9773
282 eISerAla.....LeuIleValSerLeuProTyTrpLeuAsn 293
9774 ACAAGTGGGCAATCATCGTCTGTGGTTCGCCCAAGTTTAT..... 9817
294 SerGlyTrpIleThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSe 310
9818 .....TCAGTCATCATCATCTCTCTGTCGCATAGCCTC 9852
310 rThrLeuCysGlnSerGlyIleTrpIleAsnValLeuAspIleAlaProA 327
9853 TAGCCCCCTTCACTAGAGCTAT..... 9877
327 rGlyTrSerSerPheLeuMetGlyAlaSerArgIlyPheSerSerIleAla 343
9878 .....CGTTTCCTCTTGGATGGGAACCGCT..... 9904
344 ProValIleValIProThrValSerGlyPheLeuLeuSerGlnAspProG 360
9905 ...GCCTTGATTCGCCGGGTTAAAGCCCTACTCAGCAAAATGATCCCAA 9951
360 u.....PheGlyTrpArgAsnValPhePheLeuLeu 371
9952 AGCCGCGATTTCGAGGCTTTGCCCTTCATACAGTATCTTTATCTGG 10001
371 heAlaVal 373
10002 GAGGTGTT 10009
seq_name: N_Geneseq_36:V74708
seq_documentation_block:
ID V74708 standard; DNA; 415 BP.
AC V74708:
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #397.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PN EP-786519-A2.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
PI MPI: 97-374922/35.
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
```

```
PS Claim 1: Page 1292: 3271bp: English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 415 BP; 122 A; 60 C; 88 G; 141 T;

alignment_scores:
Quality: 124.00 Length: 81
Ratio: 2.385 Gaps: 0
Percent Similarity: 64.198 Percent Identity: 30.864

alignment_block:
US-09-391-958-1 x V74708 ..

Align seg 1/1 to: V74708 from: 1 to: 415
96 TrpGlyProProGlnGluArgSerArgLeuCysSerIleAlaLeuSerG 112
112 TGGTTCCTTAATAAATGAAAAAGTAGAGCATGCAATGCTATTATTAAGCAGG 205
112 ymetLeuLeuGlyCysPheThrAlaIleLeuIleGlyGlyPheIleSerG 129
206 ATCATATTTCGACCTGATTATGACCAATAATGATTCATTGCTTTGTTA 255
129 IuThrLeuGlyTrpProPheValPheTrpIlePheGlyGlyValGlyCys 145
256 ACGCATTTAAGTCGACAGCACTATTATTAATTTTGGTCGAGTAGATTT 305
146 ValCysCysLeuLeuTrpPheValValIleTrpAspAspProValSerTy 162
306 TTAATGCGCKGATATATGGCGATATTGCCAATAAGCTTACTGAGACRACA 355
162 rProTrpIleSerThrSerGlyIlySgIuTrpIleIleSerSer 176
356 TAGMATGTTAATAGACGAGAAACGTTTCATTATGAGAAAT 398
seq_name: N_Geneseq_36:X07357
seq_documentation_block:
ID X07357 standard; DNA; 1386 BP.
AC X07357:
DT 21-MAY-1999 (first entry)
DE Staphylococcus aureus tetracycline resistance gene.
KW Tetracycline resistance; antibiotic; antibacterial; screening;
KW infection; otitis media; endocarditis; abscess; conjunctivitis;
KW epididymitis; toxic shock syndrome; sepsis; septic arthritis;
KW diagnosis; therapy; vaccine; ds.
OS Staphylococcus aureus.
PN EP-893499-A2.
PF 10-JUL-1998; 305495.
PR 23-JUL-1997; US-898976.
FT misc_feature 50..1386
FT key Location/Qualifiers
FT /tag_a /note "this nucleotide sequence is also disclosed
in the invention"
```



```
347 a[ProthrValSerGlyPhe.....LeuLeuSerGlnAspPro 359
      ::: :|::|
1250 ATTTAAAGCATTTAAATTCCAGCAATGATGCATTAAAGTTCA..... 1293
      :|::|::|
360 GluPheGlyTrrPargAsnValPhePheLeuLeuPheAlaValAsnLeuLe 376
      :|::|::|
1294 CATTTCGGT.....ATTATTTTAATTATTTAGTTTAATGAGTATTGTr 1337
      :|::|::|
376 uGlyLeuLeuPheTyrLeuIlePheGlyGluAlaAspValGlnGluTrrPA 393
      :|::|::|
1338 CGGATTAGTTTATTTCGTCATC.....TTAATCGTTGA 1372
      :|::|::|
393 lAlysGluArgLys 397
      :|::|::|
1373 CACAATCTGA AAAA 1386
```

THIS PAGE BLANK (USPTO)


```

|||||
335 TAGACAAAATGTCATCATGACATCACCATGGTAGCCATGGTCAACAGC 384
51 TheSerProGlnSerGlnLeuAspSerSerGlnValLeuProValAs 67
385 ACAAGCCCTCAATCCAGCTCAATGATTCCTGAGAGTGGCTGCTTGA 434
67 pSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysers 84
435 CTCATTGGTGGCCTAGTAAAGCCCAAGAGCTTCCTCGCAAGCTCT 484
84 erLeuLeuGlyGlyGlnPheAlaAlaLeuPheGluArgTrpGlyProGln 100
485 CAATACCTGGGGGTCACTTGGCAATTTGGGAAAGTGGGGCCCTCCACA 534
101 GluArgSerArgLeuGlySerLysAlaLeuSerGlyMetLeuLeuGly 117
535 GAACGAAGCAGACTCTGCACACTTCCTTATACAGAAATGTTACTGGAT 584
117 sPheThrAlaAlaLeuLeuLeuGlyPheLeuSerGluThrLeuGlyTrp 134
585 CTTTACTGCCATCTCATAGTGGCTTCATTAGTGAACCTTGGGTGGC 634
134 roPheValPheTrpLeuPheGlyGlyValGlyValCysCysLeuLeu 150
635 CCTTGTCTCTATATCTTGGAGGTGGCTGCTGCTGCTGCTGCTC 684
151 TrpPheValValLeuTrpAspProValSerTrpProTrpLeuSer 167
685 TGGTTGTGTGATTTATGATGACCCCGTTCTTATCCATGATTAAGCAC 734
167 rSerGluLysGluTrpLeuLeuSerLeuLysGlnValGlySers 184
735 CTCAGAAAAAGATATCATATCTCTTGAACAACAGGTGGGTCT 784
184 erLysGlnProLeuProLeuLysAlaMetLeuArgSerLeuProLeuTrp 200
785 CTAAGACAGCTCTCCATCAAAAGATATCTCATGATCTTACCATTTGG 834
201 SerLeuCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMet 217
835 TCCATATGTTAGGCGTTTACGCCATCAATGATTAAGCACAAATGCT 884
217 ValAlaTrpLeuProThrTrpLeuSerSerValTrpHisValAsnLeu 234
885 TGTATACAAACCACTTACATCAGCTGTGTATCATCATGTTAATCAGAG 934
234 sPasngLysLeuLeuSerAlaLeuProPheLeuValAlaTrpValLeu 250
935 ACAATGGACTTCTATCTGCCCTTCCTTTATTTGTCCTGGGTCAATAGC 984
251 MetValGlyGlyTrpLeuAlaAspPheLeuThrLysLysPheArgLe 267
985 AAGGTGGGGAGCTATCTGACATTTCTTAAACCAAAAGTTAGACT 1034
267 uLeuThrValArgLysLeuAlaThrLeuLeuGlySerLeuProSerSer 284
1035 CATCACTGTAGGAAATGGCACAAATTTAGGAAGTCTCCCTCTTCAG 1084
284 AlaLeuLeuValSerLeuProTrpLeuAsnSerGlyTrpLeuAlaThr 300
1085 CACTCATTTGTCTGTGCTTACCTCAATTCGGGTATATATACACCAACT 1134
301 AlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGly 317
1135 GCCTTGTGAGGCTCTCTTGGGATTAAGACAAATGTCATGTCAGGAT 1184
317 eTyTrpLeuValLeuAspLeuAlaProArgTrpSerSerPheLeuMet 334
1185 TTAATCATGCTTAAATATGCTCCAAAGGATATCCAGTTTCTCATGG 1234
334 LysAlaSerArgGlyPheSerSerLeuAlaProValLeuAlaProThrVal 350
|||||
1235 GACATCAAGAGATTTTGCAGCATAGCACCTGTCATTTGATCCACTGTC 1284
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnVal 367
1285 AGCGATTTCTTCTTAGTACGACCTGAGTTGGTGGAGGAAATCTTT 1334
367 ePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTrpLeuLeu 384
1335 CTCCTTGTGCTTGTGGCTTAACCTGTTAGGACTACTTCTACCTCATAT 1384
384 heGlyLysAlaAspValGlnGluTrpAlaLysGluArgLysLeuTrpArg 400
1385 TTGGAGAACAGATGTCATCCAAAGATGGGCTTAAGAGGAAACTCACTGT 1434
401 Leu 401
1435 TTA 1437

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-724-394A-19
seq_documentation_block:
; Sequence 19, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237/e1
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Filts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ. ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1780
; OTHER INFORMATION: /note="cDNA 22E"
US-08-724-394A-19

alignment_scores:
Quality: 2061.00 Length: 401

```

Ratio: 5.165 Gaps: 0
Percent Similarity: 99.501 Percent Identity: 99.252

alignment_block:

US-09-391-958-1 x US-08-724-394A-19 ..

Align seg 1/1 to: US-08-724-394A-19 from: 1 to: 1780

```
1 MetGlnValaspGluPheLeuIleProArgLysValProSerLeuGlySse 17
|||||
377 ATGCAAGTGTAGAGACATGATGCCAGAAAGGTCGAAGTTATGTTTC 426
17 AlaArgTyrgLyIleAlaLeuValLeuHisPheCysAsnPheThrTrt 34
|||||
427 TGCTCCGTATGGAATAGCCCTCGCTTACATTTCTGCAATTTCAACAAGA 476
34 AlaGlnAsnValIleMetAsnIleThrMetValAlaMetValAsnSer 50
|||||
477 TAGCACAATAATGTCATCATGACATCACCATGGTAGCCATGGTCAACAGC 526
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
|||||
527 ACAAGCCCTCAATCCAGCTCAATGATTCCTGAGGTGCTGCTGTTGA 576
67 PserPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSerS 84
|||||
577 CTATATTGGTGGCCCTAAGTAAAGCCCAAGAGCTTCTGCAAAAGTCTT 626
84 erLleuGlyGlyGlnPheAlaIleTrpGluArgTrpGlyProProGln 100
|||||
627 CAATACTGGGGGTGAGTTGCAATTTGGGAAAAGTGGGGCCCTCCACAA 676
101 GluArgSerArgLeuGlySerLysIleAlaLeuSerGlyMetLeuGlyGly 117
|||||
677 GAACGAAGACACTGTCACGATTCCTTATCAGGAATGTACTGGGATG 726
117 sPheThrAlaIleLeuIleGlyLysPheIleSerGluThrLeuGlyTrp 134
|||||
727 CTTTACTGCCATCTCATAGGTGGCTTCAATTAAGAAACCTTGGGTGGC 776
134 robPheValPheTyrllePheGlyGlyValGlyCysValCysLeuLeu 150
|||||
777 CTTTGGCTCTATATCTTGGAGGTGTTGGCTGCTGCTGCTGCTCTC 826
151 TrpPheValValIleTyrllePheAspProValSerTyrlleProThrIle 167
|||||
827 TGGTTTGTGTGATTTATGATGACCCCTTTCTATCCATGATGAATACAC 876
167 rSerGluLysGluTyrlleIleSerSerLeuLysGlnGlnValGlySers 184
|||||
877 CTCAGAAAAGAAATACATCATATCTCTTGAACAACAGCGGGGCTTT 926
184 erLysGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTrp 200
|||||
927 CTAAAGAGGCTCTTCCCATCAAAAGCTATGCTCAGATCTACACCATTTGG 976
201 SerLleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMetVa 217
|||||
977 TCCATATGTTAGGCTGTTTACCCATCAATGATGTTAGTACACAATGAT 1026
217 lValTyrlleProThrTyrlleSerSerValTyrlleHisValAsnIleArg 234
|||||
1027 TGTATACATACCAACTTACATCAGCTCTGTGTACCATGTTAAACATCAGAG 1076
234 sPasnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
|||||
1077 ACAATGGACTTATATGCGCCCTTATTTATGTTGGCTGGGTATATAGGC 1126
251 MetValGlyGlyTyrlleuAlaAspPheLeuLeuThrLysLysPheArgLe 267
|||||
1127 ATGGTGGAGGCTATCTGCGCAGATTTCTCTTAACCAAAAAGTTTATAGCT 1176
267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSerA 284
```

```
|||||
1177 CATCACTGTAGAGAAATTTGCCAACAATTTTAGAAGTCTCCCTCTTCAG 1226
284 lAleuIleValSerLeuProTyrlleuAsnSerGlyTyrlleThrAlaThr 300
|||||
1227 CACTCATTTGTGTCTCTTACCTCAATTCGGCTATATCAGCAACT 1276
301 AlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuGlyGlnSerGlyI 317
|||||
1277 GCTTGGCTGACGCTCTCTTGGGATTAAGCACATTTGTGATCAGGAGAT 1326
317 eTyrlleAsnValLeuAspIleAlaProArgTyrlleSerSerPheLeuMetG 334
|||||
1327 TTATATCAATGTCTTATATATGCTCCAGGATATCCAGTTTCTATG 1376
334 lYalaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
|||||
1377 GAGCATCAAGAGGATTTTCAGACATAGCACCTGATTTGATACCACTGTC 1426
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValAlp 367
|||||
1427 AGCGGATTTCTTATGTCAGGACCTGAGATTTGGGTGAGGAATGCTCTT 1476
367 ePheLeuLeuPheAlaValAsnLeuGlyLeuLeuPheTyrlleLeuIleP 384
|||||
1477 CTCTCTCTGTTTGGCCCTTAACCTGTTAGGACTACTTCTTACTCATAT 1526
384 heGlyGlnAlaAspValGlnGluTrpAlaLysGluArgLysLeuThrArg 400
|||||
1527 TTGGAGAAGCAGATGTCACAAGATGGGCTTAAGAGAGAAAAGTCACTCGT 1576
401 Leu 401
|||
1577 TTA 1579

seq_name: /cgn2_6/plodata/1/lna/5C_COMB.seq:US-08-724-394A-18
seq_documentation_block:
; Sequence 18, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Therefo
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
```

```

; TELERAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2266 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: cDNA
;   FEATURE:
;     NAME/KEY: misc.feature
;     LOCATION: 1..2266
;     OTHER INFORMATION: /note= "cDNA 22B"
US-08-724-394A-18
```

alignment_scores:

```

Quality: 932.00      Length: 471
Ratio: 3.305         Gaps: 4
Percent Similarity: 59.873      Percent Identity: 41.826
```

alignment_block:

```
US-09-391-958-1 x US-08-724-394A-18
```

```
Align seg 1/1 to: US-08-724-394A-18 from: 1 to: 2266
```

```

10 ArglyValProSerLeuGysSerAlaArgTyGlyIleAlaLeuValle 26
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 AGAAAGGTCAGATTTCTGTCATACGCTAAGGCTGGCTCTTATCAT 489
26 uHisPheCysAsnPheThrThrlleAlaIleAsnValIleMetAsnIle 43
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 GCACCTTCAAACTCCACCATGATACGCGCTGTGAGCTGTGAGCATGTG 539
43 hMetValAlaMetValAlaSerThrSerProGlnSerGlnLeuAsnAp 59
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 CGATCATGCCCATGTGAGACACACACACACAGAGCTATCTAATGCC 589
60 SerSerGluValLeuProValAspSerPheGlyGlyLeuSerLysAlaBr 76
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 TCCACTGAGGGGCTGTGACAGATGCTTCATTAACCTCAGCATATCCAT 639
76 oLysSerLeuProAlaLysSerSerIle..... 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 CAAGGAATTTGATACAAAGGCTGTGTATCATCAATGAGGCCAGAAACTC 689
85 ..... 85
690 AGGTATCATCTTTAGTCCATCACTAATGAGATATACTGACTGATC 739
86 ..... 87
740 CCAGTGTATATTAGCAGGATATTGGAGCAAAAAAATGCTGTGC 789
87 ..... 87
790 TGGTTGCTGATCTCTCCCTTCTCACCCTCTTTACACCACCTGCTGC 839
87 ..... 87
840 ACTTCGAGTGTATTGGTCATCATGTTGGAGACAGTCACAGGCGATGCC 889
88 ..... 88
890 CAGGGAATGGCATGGACAGGTCACTTATTGGGCAAGGTGGGCTCC 939
98 oProGlnIleArgSerArgLeuGysSerIleAlaLeuSerGlyMetLeuL 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
940 TCCACTTAAAGAAAGCAAGCTCACCACATTCAGAGATCAGGGTCAGCAT 989
115 euGlyCysPheThrAlaIleLeuIleGlyGlyPheIleSerGluThrLeu 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
990 TTGGATCTTCATCATCTCTGTGTGGGGGAGCTAATCTCAGAGGCTTG 1039
```

```

132 GlyTrpProPheValPheTrpIlePheGlyGlyValGlyCysAlaCys 148
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1040 AGCTGGCTTTATCTTACATCTTGTGAGCAGCGGCTGTGCTG 1089
148 sLeuLeuTrpPheValValIleTrpAspAspProValSerTyProTrpI 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1090 TCTCTATGGTTCACAGTGTATGATGATGACCCCATGATCACCCTGCA 1139
165 lSerThrSerGlyGlyTrpIleIleSerSerLeuLysGlnVal 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1140 TAAGTTAGGGGAAAGAGACATCTCTCCCTCAGCTGCTCAACAGCCC 1189
182 GLYSerSerLysGlnProLeuProIleLysAlaMetLeuArgSerLeu 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1190 AGTTCTCTGGAGAGCTGCTCCCATAAAGGAGATGTCATGCTTACC 1239
198 oIleTrpSerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSer 215
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1240 ACTTGGGCGCATTTCTGCTGGTTTTCAGCCATTTCTGTATGACACA 1289
215 hMetValAlaTrpIleProThrTrpIleSerSerValTrpHisValAsn 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1290 TCATCTTACATACCATCCACCAAGTATATCATGTACTGCTCATGTAT 1339
232 lLeuArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrp 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1340 ATCAGAGATAGGAGATTTGTCTCTCCCTGCTTTATTTGCTGCGCAG 1389
248 lIleGlyMetValGlyGlyTrpLeuAlaAspPheLeuLeuThrLysLys 264
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1390 CTGTACAAATTTTAGAGAGTCACAGCTGCGACATTTCTTTGTCTCAG 1439
265 ..PheArgLeuIleThrValArgLysIleAlaThrIleLeuGlySerLeu 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1440 TTCTCAGATTTGATCCTGTGCGAAAGCTTTTTCATCTCTTCA..... 1482
281 ProSerSerAlaLeuIleValSerLeuProTrpLeuAsnSerGlyTrp 297
1482 ..... 1482
297 eThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSerThrLeu 314
1482 ..... 1482
314 lNserGlyIleTrpIleAsnValLeuAspIleAlaProArgTyrSer 330
1483 ..... 1483
331 PheLeuMetGlyAlaSerArgLysPheSerSerIleAlaProValIle 347
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1492 TTCTTCATGGGAATCTCAAGGGATTTGGGCTCATCGCAATCATCTC 1541
347 lProThrValSerGlyPheLeuLeuSerGlnAspProGluPheGlyTrp 364
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1542 TTCCACTGCCACTGGATTCATCATGATCAGAGATTTAGCTGTGTTGA 1591
364 rGAsnValPhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPhe 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1592 GGAATGTCTTTTCTCTCTCTGCTGAGTCAACATGTTGGCTGCTGCT 1641
381 TyrLeuIlePheGlyGlyAlaAspValGlnIleTrpAlaLysGlyArg 397
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1642 TACCTCAGGTTTGGACAAAGCAAGAACTTCAAACTGGGCAAAAGAGAG 1691
397 sLeuThrArgLeu 401
1692 CCTTACCCTGCTC 1704
```

```
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-647-484-1
```

```
seq_documentation_block:
```

```
; Sequence 1, Application US/08647484
```

```
; Patent No. 5618677
```



```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-647-481-1

seq_documentation_block:
; Sequence 1, Application US/08647481
; Patent No. 5618918
; GENERAL INFORMATION:
; APPLICANT: NI, Binhui
; APPLICANT: Paul, Steven M.
; TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
; TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,481
; FILING DATE: 14-MAY-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/430,033
; FILING DATE: 27-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2716 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; : LOCATION: 461..2140
; :
; US-08-647-481-1

alignment_scores:
Quality: 479.00 Length: 397
Ratio: 2.05 Gaps: 7
Percent Similarity: 58.690 Percent Identity: 28.212

alignment_block:
US-09-391-958-1 x US-08-647-481-1 ..

Align seg 1/1 to: US-08-647-481-1 from: 1 to: 2716

12 ValProSerLeucYSerAlaArgTyrGlyIleAlaLeuValLeuThrSpa 28

```

941 ATCCCTCAGCTGCGCGGCTCCACTATGGCTGTGCATC..... 979
 28 eCyAspApherThrThIleAGlaGAsnValIleMetAsnIleThrMetV 45
 980TTCGAGAGATCTCTCAGAGCGGTGGTAGAGGGGATCACAATCC 1022
 45 aAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
 1023 CCGCGCTGC..... 1030
 62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLysSe 78
 1030 1030
 78 rLeuProAlaLysSerSerIleLeuGlyGlyGlnPheAlaIleTrpGlu 95
 1031CATGGCATCTGGAGCA 1044
 95 rGTrpGlyProProGlnGluArgSerArgLeuCysSerIleAlaLeuSer 111
 1047 AATGGCCCCCACCCTCTAGACGAGAGCGCTGGCGACGACACTTTTGT 1099
 112 GlyMetLeuLeuGlyCysPheThrAlaIleLeuIleGlyGlyPheLeSe 128
 1097 GGTTCCTATGCTGGGGCGGTGGTGGCGAGACCCCTCGCGGGGCTCTTGT 1144
 128 rGluThrLeuGlyTrpProPheValPheThrIlePheGlyValGlyC 145
 1147 GCAGTACTACAGATGAGAGCTCTGTTTCTACTGCTACAGCGCACTTCGGGA 1199
 145 yValCysCysLeuLeuTrpPheValValIleTyrAspAspProValSer 161
 1197 TCTTCTGTAACCTGTCTGGCTGCTGTCTCTACAGAGTCCCGGGGCTG 1244
 162 TyrProTrpIleSerThrSerGlyGlyGlyTyrIleIleSerSerLeuLy 178
 1247 CACCCAGCATCTCGGAGAGAGACGCCAAGTCAATC.....GA 1284
 178 sGlnGlnValGlySerSerLysGln.....ProLeu..... 188
 1285 GCAACCCATCGAGAGAGCGCGCAAACTCATGAACCCCTCAGCAAGATTAA 1334
 189ProIleLysAlaMetLeuArgSerLeuProIleTrpSerIleCys 203
 1335 GCACCTCCGCGGGGCGCTTCTACGCTATAGCAAGTATGATGCATATC 1384
 204 LeuGlyCysPheSerHisGlnTrpLeuValSerThrMetValValTyrIl 220
 1385 GGGGCGCACTTCTCGCGCAGCTGGAGAGCTTCACTGCTCCTCATGCCCA 1434
 220 eProThrTyrIleSerSerValTyrHisValAsnIleArgAspAsnGlyL 237
 1435 GCCCGCACTTCTCGAAGAAGTGTCCGCTTCAGATCAGCAAGTAGAGGCC 1484
 237 euleuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 253
 1485 TGGTGTCCGCGTCCGCCACATGTCATGATCAACATCATCTGTGCCATGCG 1534
 254 GlyTyrLeuAlaAspPheLeuLeuThrLysLysPheArgLeuIleThr.. 269
 1535 GGGCAATATCGCGACTTCTCTGGGAGACCCCGCATCATATGCCACACCA 1584
 270 ValArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaIleI 286
 1585 CTGGCCCAAGTGTAGTCACTCGCGAGAGCTTCGGCATGAGAGCCAGCTTC 1634
 286 LeuValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAlaLeu 302
 1635 TGTGTGTGTGGCTATCTCGCACTCCAAAGGGGTG...GCCATTCCTCTC 1684
 303 LeuThrLeuSerCysGlyLeuSerThrIleCysGlnSerGlyIleTyrIl 319

```

1682 CTGCTCTACCCGTGGGCTTCAGCGGCTTCCATCTCTGGTTCAACGT 1731
319 easnValleuAspIleAlaProArgTyrSerSerPheleuMeGlyAlaAs 336
1732 GAACACACTGACATAGCCCCGCGTACGACGACATCTCATGCGCATCT 1781
336 eArArgIlyPheSerSerIleAlaProValIleAlaProThrValSerGly 352
1782 CCAGAGCGCGTGGGACACACTGCGGCGATGTGCCCCATCATCGTGGG 1831
353 PheleuSerIleAspProGluPheGlyTyrPArgAsnValPhePhele 369
1832 GCCATCACTACACCAAGACTCGGAGAGTGGACATGACTGTCTCTAAAT 1881
369 uleuPheAlaValAsnleuGlyleuPheTyrIleuIlePheGly 386
1882 TGCCTCCCTGGTGCACATGAGAGTGTCTTACGGGCTTTGGCTT 1931
386 lAlaAspValGlnGlyTyrPAlaIleGlyAsnArgLysleuThr 399
1932 CTGAGAGAGACAGCGCGTGGGCGACGCTGAGAGATGAGC 1972

```

seq_name: /cgn2_6/ptodata/1/ina/5A-COMB.seq:US-08-647-481-3

```

seq_documentation_block:
: Sequence 3, Application US/08647481
: Patent No. 5618918
: GENERAL INFORMATION:
: APPLICANT: NI, Binhui
: APPLICANT: Paul, Steven M.
: TITLE OF INVENTION: HUMAN BRAIN SODIUM SODIUM DEPENDENT INORGANIC
: TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: United States of America
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/647,481
: FILING DATE: 14-MAY-1996
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/430,033
: FILING DATE: 27-APR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Gaylo, Paul J.
: REGISTRATION NUMBER: 36,808
: REFERENCE/DOCKET NUMBER: X-10006
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317) 276-0756
: TELEFAX: (317) 276-3861
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2716 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: mRNA
: US-08-647-481-3

```

```

alignment_scores:
  Quality: 479.00      Length: 397
  Ratio: 2.056        Gaps: 7
  Percent Similarity: 58.690      Percent Identity: 28.212

```

```

alignment_block:
US-09-391-958-1 x US-08-647-481-3 ..
Align seg 1/1 to: US-08-647-481-3 from: 1 to: 2716

```

```

12 ValProSerIleuCySerSerAlaArgTyrGlyIleAlaIleuValIleuHisph 28
941 AUCCGCCACACUGCCCGGUCACUAUGGUGUGUUAUC..... 979
28 ecysAsnphethrThrIleAlaGlnAsnValIleMetAsnIleThrMetcy 45
980 .....UUGUGAGAAUCCUGCAGGAGGUGUGAGAGGGGUGUACUUAAC 1022
45 AlAlaMetValAsnSerThrSerProGlnSerIleuAsnAspSerSer 61
1023 CCGCCUGC..... 1030
62 GluValIleuProValAspSerPheGlyIleuSerIleAlaProLysse 78
1030 ..... 1030
78 rIeuProAlaIysSerSerIleuGlyGlyIlePheAlaIleTyrGluA 95
1031 .....CAUGGAGUCCGAGCA 1046
95 rGTrGlyProProGlnGluArgSerArgLeuCySerIleAlaIleuSer 111
1047 AAUUGGCCCCACCUCUAGAAGGAGGCGCCUGGCGCAGCAGACUUCUUGU 1096
112 GlyMetLeuIleuGlySerPheThrAlaIleuIleGlyIlePheIle 128
1097 GGUUCCUAUGCGUGGGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1146
128 rGluThrIleuGlyTyrProPheValPheTyrIlePheGlyIleValGly 145
1147 GCAGUACUACAGUAGAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAG 1196
145 ysValIleuCySerIleuThrPheValIleTyrAspAspProValSer 161
1197 UCUUCUGUACUUCUUCUGUGUGUGUGUGUGUGUGUGUGUGUGUGUG 1246
162 TyrTrpIleSerThrSerGluGlyGlyTyrIleIleSerSerIleu 178
1247 CACCCAGCAUCCUGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1284
178 sGlnGlnValGlySerSerIleGln.....Proleu..... 188
1285 GGAGCGCAUCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1334
189 .....ProIleuValMetIleuArgSerIleuProIleTyrSerIleCy 203
1335 GCACUCCUGGCGGCGGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUG 1384
204 LeuGlyCyPheSerHisGlnThrIleuValSerThrIleMetValIle 220
1385 GUGGCGCAUCCUGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1434
220 eProThrTyrIleSerSerValTyrHisValAsnIleArgAspAsnGly 237
1435 GCCCGCAUCCUUCGAAAGAGUUGUGUGUGUGUGUGUGUGUGUGUGUG 1484
237 euleuSerAlaIleuProPheIleValAlaTyrValIleGlyMetValGly 253
1485 UGUGUGUGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1534
254 GlyTyrIleuAlaAspPheleuThrIleuTyrIleuTyrIleuTyr 269
1535 GGCGAGAUCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1584
270 .ValArgLysIleAlaThrIleLeuGlySerIleuProSerSerAlaIle 286
1585 CGUGGCAAGUGUAGAACUGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1634

```



```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
US-08-430-033A-3

Alignment_scores:
      Quality: 479.00      Length: 397
      Ratio: 2.056      Gaps: 7
Percent Similarity: 58.690      Percent Identity: 28.212

Alignment_block:
US-09-391-958-1 x US-08-430-033A-3      ..

Align seg 1/1 to: US-08-430-033A-3 from: 1 to: 2716

12 ValProSerLeuGlySerAlaArgTyrGlyTlleAlaLeuValLeuHisHsp 28
   :::::::::::::::::::::
941 AUCCCCUCCAGUCUCCGCGCCACUAGGCGUGUC..... 979

28 ecysAsnpherThrTlleAlaGlnAsnValIleMetAsnIleThrMetV 45
   ::::: ||| ::::::::::: ::::::::::
980 .....UUCGUGAGAUCCUGACGGGUGUGUGAGAGGGGUCACAUAC 10222

45 aAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
   |||
1023 CCGCCUGC..... 10330

62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLys 78
1030 ..... 10330

78 rLeuProAlaLysSerSerTlleGlyGlyGlnPheAlaTlleTrpGlnA 95
   :::::::::::::::::::::
1031 .....CAUGGAGUCCUGAGACA 1046

95 rGTrpGlyProProGlnGluArgSerArgLeuGlySerTlleAlaLeuSer 111
   :::::::::::::::::::::
1047 AAUGGGCCCCACCCUUGAGAACGAGUGCCUGCGACGACAGACCCUUDUGU 1096

112 GlyMetLeuLeuGlyCysPheThrAlaIleLeuIleGlyGlyPheIle 128
   ||| ::::::::::: ::::::::::: ::::
1097 GGUUCUUAUGUCUGGGCGGUGUGCGGAGUCCUUGCGCGGGUCCUGU 1146

128 rGluThrLeuGlyTyrProPheValPheTyrTllePheGlyGlyValGly 145
   ::::: ::::::::::: ::::::::::: |||
1147 GCAGAACCCAGGADGAGCUCUGUUUUUACGUCUACGCGACGCUUGCGGA 1196

145 ySValCysCysLeuLeuTrpPheValValIleTyrAspAspProValSer 161
   ||:::::::::::::::::::
1197 UCUUCUGGUACUUCUUCUGCGUCUGUCUCCUACAGAUCCCCCGGCGUG 1246

162 TyrProTrrPleSerThrSerGlnLysGlyTrrIleIleSerSerLeu 178
   ::::: ||| ::::: ::::::::::: ||| ::
1247 CACCCCAACAUUCUGAGAGGAGCGCAGUACUAC.....GA 1284

178 sGlnIleValGlySerSerLysGln.....ProLeu..... 188
   ::::: ::::::::::: |||
1285 GGACGCCACUCCGAGAGAGCGGAAACUACUAGAACCCUCCACGAGAUUUA 1334

189 .....ProIleLysAlaMetLeuArgSerLeuProIleTrpSerIleCys 203
   ||| ::::: ::::: ||::::::::: |||
1335 GCACUCCCGGCGGCGCUUUCACGUCUACUAGCCAGAGUUAUGGCACUAC 1384

204 LeuGlyCysPheSerHisGlnTrpLeuValSerThrMetValValTrr 220
   ::::::::::: ||| :::::
1385 GUUGGCAACUUCUGCGAGUGGAGCGUUCUACUUGUCUCCUUAUCCCA 1434

220 eProThrTrrIleSerSerValTrrHisValAsnIleArgAspAsnGly 237
   ||| ::::::::::: ||::: ::::: |||
1435 GCCGAGUACUUCGAAAGAUUGUUGCGUCCGACAGUACAGCAAGUUGAGCC 1484

```

```

237 euleuseralaleupropheilevalatrypallleglymetvalgly 253
|||||
1485 UGGUGUGCGGGCGGCCCGCCCGUGGUGCAUGAACACACACGUGGCCAUGGCG 1534
|||||
254 G1YTYrleualaaspheuleuleuthrlyslsypheargleuilethr. 269
|||
1535 GGGCAGAGGCGGCGACUCCUGCGGAGCGCGCGCAUACUAGCCACCACCAA 1584
|||||
270 .ValarglyllealathrilleuuglyserleuproserSerlaleu 286
|||||
1585 CGGGCGCAUGUUGAUGACGCGGAGGCGUUGCGCAUGGAGCCACGCGUG 1634
|||||
286 levalserleuprotyrleuanserglytyrillethralthalaleu 302
|||||
1635 UGUUGUGUGGCGGUGACUCCGACUCCCAAGGGCGUG..GCCAUCCUCC 1681
|||||
303 leuthrleuSerCysglyleuSerthrleuGslInserglyleuYrll 319
|||||
1682 CUGGUGCCAGCGGUGGCGUUCGACGCGGUGGUGGCAUCCUGGUGUACAGU 1731
|||||
319 eAsnValleuAspIlealaproargtyrSerSerPheuleumetglyalas 336
|||||
1732 GAACGACCUUGGACAUAGCCCGCGCUACGCGCACGACAUCCUACUGGGCAUCU 1781
|||||
336 eArGglyPheSerSerIlealaproValIleValProthrValSergly 352
|||||
1782 CCAACGCGCGGCGGACACUGUGGCGCAUGGUGUCCCGCCAUCAUCCUGGCG 1831
|||||
353 PheleuSerleuAspProgluPheglytyrParGAsnValPhePhe 369
|||||
1832 GCCAUGACUAGCACAGACUCCGCGGAGAGUGGCAUGAGUGUCCUUAU 1881
|||||
369 uLeuPheAlaValAsnleuGlyleuLeuPheTyrlleuIlePhegly 386
|||||
1882 UGCCUCCUGGUGCAUAGGAGUGGUGAUCCUACGCGGUGUCCUUCU 1931
|||||
386 luAlaAspValGlnGluTrpAlaLysGluArgLysleuThr 399
|||||
1932 CUGAGAGAGAGACGCGGUGGCGACAGCCUGAGGAGAUAGAC 1972
|||||
seq_name: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:PCT-US96-05792-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9605792
; GENERAL INFORMATION:
; APPLICANT: NI, Binhui
; APPLICANT: Paul, Steven M.
; TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
; TITLE OF INVENTION: PHOSPHATE COTRANSPORTER
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05792
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/430,033
; FILING DATE: April 27, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10006

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2716 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 461..2143
; PCT-US96-05792-1

alignment_scores:
Quality: 479.00 Length: 397
Ratio: 2.056 Gaps: 7
Percent Similarity: 58.690 Percent Identity: 28.212

alignment_block:
US-09-391-958-1 x PCT-US96-05792-1 ..

Align seg 1/1 to: PCT-US96-05792-1 from: 1 to: 2716

12 ValProSerLeuCysSerAlaArgTyrGlyIleAlaLeuValLeuHisPh 28
:::|||||
941 ATCCCTCAGCTGCCCGGCTGCACATGCTGTGCATC..... 979
28 eCysAsnPheThrThrIleAlaGlnAsnValIleMetAsnIleThrMetV 45
|||||
960 .....TTCTGAGAGATCTTCGACGGGCTTGAGAGGGGTCACATACC 1022
|||||
45 AlaIleMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
|||
1023 CCGCCTGC..... 1030
62 GluValLeuProValAspSerPheGlyLeuSerLysAlaProLysSe 78
|||||
1030 ..... 1030
78 rLeuProAlaLysSerSerIleLeuGlyGlnPheAlaIleTrpGluA 95
|||||
1031 .....CATGGATCTGGAGCA 1046
95 rGTrGlyProProGlnGluArgSerArgLeuCysSerIleAlaLeuSer 111
|||||
1047 AATGGGCCCAACCTTAGAGGAGATGCGTGGCGACAGACGCTTTGT 1096
|||||
112 GlyMetLeuLeuGlyCysPheThrAlaIleLeuIleGlyGlyPheIleSe 128
|||||
1097 GGTTCCTATGCTGGGGCGGTGCTCGCATGCCCGCGGGGCTCTTGT 1146
|||||
128 rGluThrLeuGlyTrpProPheValPheTyrllePheGlyGlyValGlyC 145
|||||
1147 GCACTACTCAGATGAGCTCTGTTTCTACGCTACGCGACGCTTGGA 1196
|||||
145 ySValCysLeuLeuTrpPheValIleTyraSPAspProValSer 161
|||||
1197 TCTTCTGTACTCTTTCGCGTCTGCTCTACGATCCCGCGGCTG 1246
|||||
162 TyrProTrpIleSerThrSerGluysclutrylleIleSerSerleuLy 178
|||||
1247 CACCCACACATCTCGAGAGAGAGCGCAAGTACATC.....GA 1284
|||||
178 sGlnGlnValGlySerSerLysGln.....ProLeu..... 188
|||||
1285 GACGCCCATCGGAGAGCGGCAAACTCATGACCCCTCCAGCAAGTTTA 1334
|||||
189 .....ProIleLysAlaMetLeuArgSerLeuProIleTrpSerIleCys 203
|||||
1335 GCACCTCCCTGGGGCGGCTTCTTACGCTCATGCCAGTATCATCATCATC 1384

```

```

204 LeuGLyCySeSerHisGlnTrpLeuValSerThrMetValValTyr11 220
      ::::: ||| ::::: ||| ::::: |||
1385 GTGGCCAACTTGTGGCGAGCTGAGCTTCTACCTGCTGCTATCTCCCA 1334
      ||| ::::: ||| ::::: |||
220 eProrhTrpTrpLeuSerValTyrHisValAsnIleArgAspAsnGlyL 237
      ||| ::::: ||| ::::: |||
1435 GCCCGACTCTTCGAGAGAGTGTTCGGCTTCGATCCAGCAAGTGGCC 1484
      ||| ::::: ||| ::::: |||
237 euleuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 253
      ||| ::::: ||| ::::: |||
1485 TGTGTGTCGGCGCTGCCCCACCTGCTGATGACATCTTCGTCCTTC 1534
      ||| ::::: ||| ::::: |||
254 GlyTyrLeuAlaAspPheLeuLeuThrTyrLysPheArgLeuIleThr 269
      ||| ::::: ||| ::::: |||
1535 GGCCAGATGCGGAGCTTCTGCGGAGCCGCCGATCATGTCACACCA 1584
      ||| ::::: ||| ::::: |||
270 ValArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu 286
      ||| ::::: ||| ::::: |||
1585 CGTGGCGAGTTGATGACTGCGGAGCTTCGGCATGAGAACCCGCTGC 1634
      ||| ::::: ||| ::::: |||
286 LeuAlSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAlaLeu 302
      ||| ::::: ||| ::::: |||
1635 TGTGTGTCGCTGCGCTACTGCGACCTCCAAGGGCGTG...GCCATCTCTTC 1681
      ||| ::::: ||| ::::: |||
303 LeuThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGlyIleTyr11 319
      ||| ::::: ||| ::::: |||
1682 CTGGTCTTACCGGTGGGCTTCAAGCGCTTCGCCATCTCTGGGTTCAACGT 1731
      ||| ::::: ||| ::::: |||
319 eAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetGlyAla 336
      ||| ::::: ||| ::::: |||
1732 GAACCACTGAGACATAGCCCGCTACCGCCAGCATCTCATGGGATCT 1781
      ||| ::::: ||| ::::: |||
336 eArgGlyPheSerSerIleAlaProValIleValProThrValSerGly 352
      ||| ::::: ||| ::::: |||
1782 CCAACGGCTGGCGACACTGTGGGATGTGTGCCCATCATCGTGGGG 1831
      ||| ::::: ||| ::::: |||
353 PheLeuLeuSerGlnAspProGlnPheGlyTyrPArgAsnValPhePheLe 369
      ||| ::::: ||| ::::: |||
1832 GCCATGACTACACACAGACTCGGAGAGTGGAGTACGTGTTCTTAAT 1881
      ||| ::::: ||| ::::: |||
369 uLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuIlePheGly 386
      ||| ::::: ||| ::::: |||
1882 TGCCTCCCTGCTGACTATGAGGTGTCTTACGGGGCTTTTGGCTT 1931
      ||| ::::: ||| ::::: |||
386 IuAlaAspValGlnGlnTrpAlaLysGlnArgLysLeuThr 399
      ||| ::::: ||| ::::: |||
1932 CTGAGAGAGAGAGCGCGTGGCGAGAGCCCTGAGAGATGAGC 1972
      ||| ::::: ||| ::::: |||
seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-05792-3

seq_documentation_block:
: Sequence 3, Application PC/TUS9605792
: GENERAL INFORMATION:
: APPLICANT: N1, Binhui
: APPLICANT: Paul, Steven M.
: TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: United States of America
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/05792

```

```

: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/430,033
: FILING DATE: April 27, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Gaylo, Paul J.
: REGISTRATION NUMBER: 36,808
: REFERENCE/DOCKET NUMBER: X-10006
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317) 276-0756
: TELEFAX: (317) 276-3861
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2716 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: RNA
: PCT-US96-05792-3

alignment_scores:
      Quality: 479.00      Length: 397
      Ratio: 2.056      Gaps: 7
      Percent Similarity: 58.690      Percent Identity: 28.212

alignment_block:
US-09-391-958-1 x PCT-US96-05792-3 ..

Align seg 1/1 to: PCT-US96-05792-3 from: 1 to: 2716

12 ValProSerLeuCysSerAlaArgTyrGlyIleAlaLeuValLeuHisph 28
      ::::: ||| ::::: ||| ::::: |||
941 AUCCCCUACAGUCGCCGCGUCACUAGUCUGUCUACU..... 979
      ||| ::::: ||| ::::: |||
28 eCysAsnPheThrThrIleAlaGlnAsnValIleMetAsnIleThrMetV 45
      ||| ::::: ||| ::::: |||
980 .....UUCGUGAGAGUUCUGCAGGCGUGUGUAGAGGGGUCACUACU 1022
      ||| ::::: ||| ::::: |||
45 AlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
      ||| ::::: ||| ::::: |||
1023 CGCCCGC..... 1030
      ||| ::::: ||| ::::: |||
62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLys 78
      ||| ::::: ||| ::::: |||
1030 ..... 1030
      ||| ::::: ||| ::::: |||
78 rLeuProAlaLysSerSerIleLeuGlyGlnPheAlaIleTrpGluA 95
      ||| ::::: ||| ::::: |||
1031 .....CAUGGAGUCUGAGCA 1046
      ||| ::::: ||| ::::: |||
95 rGTrpGlyProProGlnGlnArgSerArgLeuCysSerIleAlaLeuSer 111
      ||| ::::: ||| ::::: |||
1047 AAUGGGCCCCACCCUAGAGAGGCGCCUGGCGACGACACCCUUDUUG 1096
      ||| ::::: ||| ::::: |||
112 GlyMetLeuLeuGlyCysPheThrAlaIleLeuIleGlyGlyPheLeu 128
      ||| ::::: ||| ::::: |||
1097 GUUCCUAGUCUGGGGCGUGGCGGAGUCCUUGCGCGGUGCCUUGU 1146
      ||| ::::: ||| ::::: |||
128 rGlnThrLeuGlyTyrProPheValPheTyrIlePheGlyGlyValGlyC 145
      ||| ::::: ||| ::::: |||
1147 GCAGUUCUAGAGUGAGCUCUGUUCUACGUCUACGACACUUCGGA 1196
      ||| ::::: ||| ::::: |||
145 yValAlaCysCysLeuLeuTrpPheValValIleTyrAspAspProValSer 161
      ||| ::::: ||| ::::: |||
1197 UCUCUCUGUACUCUUCUUGCUGCUCUGUCUACAGAGUCCCGCGCUG 1246
      ||| ::::: ||| ::::: |||
162 TyrProTrpPheSerThrSerGlyGlyGlyTyrIleLeuSerSerLeu 178
      ||| ::::: ||| ::::: |||
1247 CACCCAGCAUCUCGAGAGAGGAGCGCAAGUACU.....GA 1284
      ||| ::::: ||| ::::: |||
178 sGlnGlnValGlySerSerLysGln.....ProLeu..... 188

```

```
1285 GAGCGCAUCGAGAGAGCGCAACUAGAACCCGCUACGAGUUUA 1334
      ::::: |||||:::
189 ....ProilelysaMetleuArSerleuProiletpSerleCy3 203
      ||| ::::: |||||:::
1335 GCACUCCGCGCGCCGCUUCACGCUUAGGAGUCUAGGCAUUC 1384
      ::::: ||| :::::
204 leuGLCyphSerHisGIntRpleuValSerThrMetValYrIl 220
      ::::: ||| :::::
1385 GUGGCGCAUCUCCGCGCACUGGAGGUCUCCUGGUCUCCCA 1434
      ||| |||||:::
220 eProthTyRleSerSerValTyRHisValAsnleArGAspAsnlyl 237
      ||| |||||:::
1435 GCGCGACUACUUCGAGAAAGUUGCGCUUCGAGAUACGAGUAGGC 1484
      |||||:::
237 euleSerAlaleuProPheIleValAlatrPValIleGlyMetValGly 253
      |||||:::
1485 UGUUGUCCGCGCGCCACCUUGUCUAGCACAUCCGUGCCCAUCGGC 1534
      ||| :::::
254 GlyTyRleuAlaAspPheleuLeuThrLysLysPheArgleuIleThr.. 269
      ||| ::|||::|||
1535 GGCAGAUCCGGAGACUCCGCGAGCGCGCAUCAGUCCACACCA 1584
      |||||:::
270 .ValArgLysIleAlaThrIleleuGlySerleuProSerSerAlaleu 286
      |||||:::
1585 CGUGCGCAUUGAUAGAACUGCGGAGCUUCGCGCAUAGCACCGCUGC 1634
      |||||:::
286 leValSerleuProTyRleuAsnSerGlyTyRleThralaThralaLeu 302
      ::::: ||| ::|||
1635 UGUUGUGGCGCGCUCCACUCCCAAGGCGUG...GCCAUCCUCCUG 1681
      |||||:::
303 leuThrleuSerCysglyleuSerThrleuGlylnSerIleTyRyl 319
      |||||:::
1682 CUGGUCUCCAGCCGUGGCUCCGCGCUCCGCAUCCUGGUCUCCAG 1731
      |||||:::
319 easnValleuAspIleAlaProArGtyrSerSerPheleuMetCylAlas 336
      |||||:::
1732 GAACCAACUGGACAUAGCCCGCGCUACGCGCAGAUCCUACUGGCAUUC 1781
      |||||:::
336 eArGglyPheSerSerIleAlaProValIleValProThrValSerGly 352
      |||||:::
1782 CCAACGCGCGGCGCACUCUGCGGCUCCGCCAUCUCCGUGGCGG 1831
      |||||:::
353 PheleuSerGlnAspProGluPheGlyTyrArgAsnValPhePhele 369
      ::::: |||||:::
1832 GCCAUGGCUAGCAACAGACUGGAGAGUGGCAUGAGUGUCCAU 1881
      |||||:::
369 uleuPheAlaValAsnleuLeuGlyleuLeuPheTyRleuIlePheGly 386
      |||||:::
1882 UGCUCUCCUGGUCACUAGGAGUGUCCAUCCUACGCGGUCUUCGU 1931
      |||||:::
386 luAlaAspValGlnGluTyrPalalysGluArgLysleuThr 399
      ::::: ||| ::|||
1932 CUGGAGAGAGCGCGCGGAGAGCCUUGAGAGAGAGAGC 1972
      |||||:::
seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-805-118-5
```

```
seq_documentation_block:
; Sequence 5, Application US/08805118
; Patent No. 5985604
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
; TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,118
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0221 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAITVOT2
CLONE: 754412
US-08-805-118-5

alignment_scores:
Quality: 433.00 Length: 85
Ratio: 5.094 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x US-08-805-118-5 ..

Align seq 1/1 to: US-08-805-118-5 from: 1 to: 272
317 11eTyRleAsnValleuAspIleAlaProArGtyrSerSerPheleuMe 333
|||||
1 ATTATATCAATGTCCTTAGATATTGCTCCAAAGATATTCACATTTCAT 50
|||||
333 tGlyAlSerArGlyPheSerSerIleAlaProValIleValProThrv 350
|||||
51 GCGAGCATCAAGAGATTTTCGACATAGACACCTGTCATTGTACCACTG 100
|||||
350 a1SerGlyPheleuLeuSerGlnAspProGluPheGlyTyrPArgAsnVal 366
|||||
101 TCACTTGATTTCTTCTTAGTCAGACCCCTGAGTTGGGTGAGGAATGTC 150
|||||
367 PhePheleuLeuPheAlaValAsnleuLeuGlyleuLeuPheTyRleuIl 383
|||||
151 TTCTTCTTGCTGTTGGCGTTAACCTGTTAGGACTCTTCTACCTCAT 200
|||||
383 ePheGlyluAlaAspValGlnGluTyrPalalysGluArgLysleuThra 400
|||||
201 ATTGTGAGAGACAGATGTCACAGAAATGGGCTAAAGAGAGAAATCACTC 250
|||||
400 rGleu 401
|||||
251 GTTAA 255

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-724-394A-20

seq_documentation_block:
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
```

137	phenylleuPhe.....glycylalglycylalcyScysleuLeu	151
22330	TTTTATCCCTCTCTCCAGAGTGCACACTGGCTGTGCTGTCTCCTTANG	22281
151	pheValValIleIleIleTyrAspAspProValSerIleTyrProIleSerIle	168
22280	GTTCCACGTGATTATGTGATGACGCCCATGCATGCATCACCCTGACATAAGTGTTA	22231
168	erGluIuIuSerIleIleIleSerSerLeuIuIuValGly.....	182
22230	GGGAAAGAGAGACACATCCTGTCTCTCACTGGGCTCACACAGTACAGTGACA	22181
183Se	183
22180	CCTGTACTCTGGCCCATGCAGAGCTCTAAGGCCAGGGTGTGATCTC	22131
183	rSerIuSGLn...Pro.....	187

22130	CTGTGAGGACACCATCTTGGCTGTCTTAATCTATGCTGATTAAGTCT	22081LeuProIIeIysAlaMetL	194
198	194	22080
22080	TTCCTTTCAAGCCAGTTCCTCTGGACGAGCTGTGCCCCATTAAGGGCATGG	22031	22031
194	euarSerLeuProIIerPserIIeCysLeuGIcysPheSerHisGln	210	22030
22030	TGCATGCGCTACACACTTTGGGCATTTTCCCTGGGTTTTTTCAGCCATTTC	21981	22030
211	TTPLeuValSerThermetalValIYrIIeProthIYrIIeSerSera	227	211
21980	TGGTGTGGACCATCATCTCAACATACCTACCAAGATATACGATCTCT	21931	21980
227	IYrHisValasnIIeArgAsp.....	234	227
21930	GCTCCATGTTCATCATCAGAGATGTGAGTTACTCTCATACTTACGAA	21881	21930
234	234	234
21880	AATGATTAATGTAAATAGAGAGAAACAGTCTGTGTACCTATTACATTCT	21831	21880
234	234	234
21830	GGCTTTACATAAACCATTAATTAACTTACACTTACAAATGACTTGAGAGAG	21781	21830
234	234	234
21780	CATTGTATTAATTCCTTTTCACAGATGTGGAAACAGACACTTAGAGGT	21731	21780
234	234	234
21730	GAGATACTTGCCCAAGTGTGCACAATACTAAGTAGAGCTGCTGCAG	21681	21730
234	234	234
21680	CATCATATTTCTTAACCACTATGCTATACTACACACCAAGCTGATCCAA	21631	21680
234	234	234
21630	AGCTTCTTTAGAAATAATATTGCTGGGCCAGCATGTGTCAATGCTG	21581	21630
234	234	234
21580	TAATTCCAGCACTTGGAGAGGCCAGGACGACAGATCATGAGTCAAGAA	21531	21580
234	234	234
21530	TGCAGACCAAGCTTGACCAATATGTTTACTAAATATCATCTACTAAAAA	21481	21530
234	234	234
21480	TACAAAAATTAGCCAGGTGTGTGGACAGCAGCTGTAAATCCAGCTATTTC	21431	21480
234	234	234
21430	AGAGAGCTGAGACAGAGAAATCGTGTGAACCCAGAGAGGTGAGGTTGCAT	21381	21430
234	234	234
21380	TGAGCCAAGATCATGCACCTGCACTCCAGCCTGGGCGACAGAGTAAAGCT	21331	21380
234	234	234
21330	CCGTTTCAAAAAACAAAAACCCCAAGAAATTATATTGCTTTTATCTGGAG	21281	21330
234	234	234
21280	CCCAAGATGATCAGACTTCTGGCCCTCTTATCTGAGACAGTCTTTTA	21231	21280
234	234	234
21230	GTGTGAAAAAGGATGCTAATTTTCCCCCAACAAACCCACAGTATCATGGG	21181	21230

```
234 ..... 234
21180 GGTAAAGTAAATGCTGTGTGTACTGACAAATTTGGTGCACGTA 21131
235 ..... Asn..GlyLeuLeus 239
21130 TCTTAACTACTCTGTATAAATCTTCCTTCACAGAGGAGTCTGT 21081
239 ealalaleupropheilevalalatrivalleglmetvalglylytyr 255
21080 CTTCCCTGCTTTATGCTGCTGCAAGCTGTACAAATTTAGAGAGTCAG 21031
256 leualasphleuleuthrlylsys...pheargleuileuprvalar 271
21030 CTGGCAATTTCTTTGTCAGGAATCTTCAGATTGACTGTGCG 20981
271 glysilalathrileuuglyser.leu..... 280
20980 AAGCTCTTTTCATCTCTGTGTAGATTAAGCGTGGCCCATTTAAC 20931
280 ..... 280
20930 AATCCCTTTCTGCACATGCTCTCAGAGGTTCCCTGACAGCATGTCTC 20881
281 ..... Proser.Seralaleuilevalserleuprottyr 291
20880 ATTGCCAGAGGCTCTCTTCATCAATATGTGCTGTGGCCCTCTTT 20831
292 leuansersglytyrilethrilaralaleuethrleuusercysgl 308
20830 GTGGCCCTCAGTTACGTATACCATATTTTGTGATACTATTTCTCG 20781
308 yleuserthrleucysglinsersglylytyrileasvalleuaspillea 325
20780 GACCAAGTAACTATGTACTCAGGGTTTATCATCAACACCTTAGATATCG 20731
325 laproargtyrserserpheleu 332
20730 CCCCCAGTAAAGACTCTACCTG 20708
seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:us-08-724-394A-21
seq_documentation_block:
; Sequence 21, Application US/08724394A
; Patent NO. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TYPE OF INVENTION: Sequences and Antibodies thereo
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
```

```
ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 01/957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note="HLA-H.CONTIG"
US-08-724-394A-21
```

alignment_scores:

Quality:	297.50	Length:	543
Ratio:	1.848	Gaps:	10
Percent Similarity:	29.650	Percent Identity:	20.994

alignment_block:

US-09-391-958-1 x US-08-724-394A-21/rev ..

Align seg 1/1 to reverse of: US-08-724-394A-21 from: 1 to: 246240

```
137 PheTyrllePhe.....GlyGlyValGlyCysValCysCysLeuLeuTr 151
22330 TTTTATCCCTCTTCCAGGAGGACGTGCTGTCTGCTCTCTCTATG 22281
151 PheValValleTyrraspaaprovalserTyrrProTrilleserThrs 168
22280 GTTCACAGTGTATTATGATGACCCCATGATCACCCTGATTAAGTGTTA 22231
168 erglutygslytyrilleleserleuysglnvalgly..... 182
22230 GGGAAAGGAGACATCTCTGCTCTACAGCGCTCAACAGGTACAGTGCACA 22181
183 .....Se 183
22180 CTTGTACTGTGGCCCATGCAGAGGTCTTAGGGCAGAGGTGGATCTC 22131
183 rserlysgln..Pro..... 187
22130 CTCTGAGAGGCACCATCTTGGCTGCTCTAATATCATGATGATTAGATCT 22081
188 .....LeuProilleyslameLl 194
22080 TTTCTTTCAGCCAGCTTCTCTGAGCAGAGCTGTCCCATTAAGGCGATGG 22031
194 euargserleuproilletrpserillecysleuglyCysPheSerHsGln 210
22030 TCACATGCTTACACTTGGGCAATTTTCCGCGGTTTTCAGCCATTTTC 21981
211 TrpleuValserThrmelvalValTyrlleprohTyrrileserSeVa 227
21980 TGGTTGTGCACCATATCTTAACATACCTTACCAAGATATATCATGACTCT 21931
227 lTyrrHisValasnilleargasp..... 234
21930 GTCCTCATGTTAACATCAGAGATGTGAGTTTACTCTCATATTCTTACGAA 21881
234 ..... 234
21880 AATGATAATGTAATTAAGAGAAACAGTCTGTGTACTACTATTAATCT 21831
234 ..... 234
```

21830 GCGTTACATATATACCATTAATTTACCTTCACATGACCTTGAGAGG 21781
 234 234
 21780 CATGTATATATTCCTTCACAGATGTGAAACAGACACTTAGAGT 21731
 234 234
 21730 GAGATACCTTGCCCGAGTTGCACATATCTAAGTATAGAGCTGCTGAC 21681
 234 234
 21680 CATCCATATCTTACCATATGCTATCTACACACACAGCTGATTCGA 21631
 234 234
 21630 AGCTTCTTTAGAAATATATGTGCTGGCCGAGCATGTGTGCTGCTG 21581
 234 234
 21580 TAAATCCAGCATTTGGGAGCCGAGCAGACATCATGAGTCAAGAA 21531
 234 234
 21530 TGCAGACAGCGCTGACCAATATGTTTACTAATATCATCTACTAAAA 21481
 234 234
 21480 TACAAAAATTAGCCAGGTGTGTGGGAGGACCTGTATCCAGCTATTC 21431
 234 234
 21430 AGGAGCTGAGACAGAGAAATCCCTTGAACCCAGAGGTGAGTTGCAT 21381
 234 234
 21380 TGAGCCAAATCTGCGACTGCGCTCAGCTGGGAGAGAGTAAGACT 21331
 234 234
 21330 CCGTTTCAAAAACAAAAAACCAAGAAATATATGTTTATCTGGAG 21281
 234 234
 21280 CCGAGATGATGAGCTGCTGGCCCTTATCTGAGACAGTGTCTTTA 21231
 234 234
 21230 GTGTGAAAAAGATGCTATTTTCCCCCAACCAACCATATCATGGG 21181
 234 234
 21180 GGTAAATTAATGCTGTCTGTATCTGACAAATTTGGTGTAACTGA 21131
 235 235
 21130 TCTCTATTAATCTGTAATAACTTCTCTCTGAGAGTGTGCTGT 21081
 239 239
 239 239
 21080 CCGCTGCTGCTTTATGCTGCTGCAAGTGTACATTTTAGAGAGTCA 21031
 236 236
 21030 CTGGCAGATTTCTTTTCCAGGAATCTTCAGATGATGATGAGTGG 20981
 271 271
 271 271
 20980 AAAGCTTTTCACTCTTGTGAAGATAAGCGTGTGGCCCATTTAAC 20931
 280 280
 20930 AATCCCTTTTTCGACATGTGTCTCAGAGGTTCCCTGACAGATGTCTC 20881

281Proser.SerAlaLeuIleValSerLeuProTyr 291
 20880 ATGGCCAGGCTCCCTTCATCATATATGCTGTGGCCCTGCTT 20831
 292 LeuAsnSerGlyTyrIleThrAlaThrAlaLeuThrLeuSerGly 308
 20830 GTGGCCCTCAGTATGATGATACCATATTTGCTGATCTATTCCTG 20781
 308 yLeuSerThrLeuGlyGlnSerGlyIleTyrIleAsnValLeuAspIle 325
 20780 GACCAGTAACCTATGTGACTCAGGTTTATCATCAACACTTAGATATG 20731
 325 LabProArgTyrSerSerPheLeu 332
 20730 CCCCAGGTAAAGCTTACCTG 20708
 seq_name: /cgn2_6/plodata/1/lna/5C_COMB.seq:US-08-724-394A-22
 seq_documentation_block:
 ; Sequence 22, Application US/08724394A
 ; Patent No. 5872237
 ; GENERAL INFORMATION:
 ; APPLICANT: Feder, John N.
 ; APPLICANT: Krommal, Gregory S.
 ; APPLICANT: Lauer, Peter M.
 ; APPLICANT: Ruddy, David A.
 ; APPLICANT: Thomas, Winston
 ; APPLICANT: Tsuchihashi, Zenta
 ; APPLICANT: Wolff, Roger K.
 ; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/724,394A
 ; FILING DATE: 01-OCT-1996
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pitts, Renee A.
 ; REGISTRATION NUMBER: 35,136
 ; REFERENCE/DOCKET NUMBER: 017957-000100
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-576-0200
 ; TELEFAX: 415-576-0300
 ; INFORMATION FOR SEQ ID NO: 22:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 246240 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..246240
 ; OTHER INFORMATION: /note= "HLA-H.CONTIG"
 ; US-08-724-394A-22
 alignment_scores:
 Quality: 297.50 Length: 543
 Ratio: 1.848 Gaps: 10

Percent Similarity: 29.650 Percent Identity: 20.994

US-09-391-958-1 x US-08-724-394A-22/rev

Align seg 1/1 to reverse of: US-08-724-394A-22 from: 1 to: 246240

137	PheTyrIlePhe.....GlyGlyValGlyCysValGlyCysLeuLeuTr	151
22330	TTTTATCCCTTTTCCCAGGTACACACTGGCGTGTCGTGCTGTCTTAATG	22281
151	pPheValValIleTyrAspAspProValSerTyrProTriPleserThrs	168
22880	GTTCAcAGtAATTATtANGAtSACCcCAGtGCATtAcGCCgGtCATtAGtGTtA	22231
168	eRGlulysGIuTrrIleIleSerSerleuLysGIInValGly.....	182
22230	GGGAAAGGAGCACATCTGTCTCACTGGCTCAACAGGTACAGTGACA	22181
183Se	183
22180	CCTTGTACTGTGGCCCAATGCAGAGGCTTAGGGCAGGCTGTGATCTC	22131
183	rSerLysGIIn..Pro.....	187
22130	CTCTAGAGGACACATCTTGCTGTCTAATtACtATGCtGtATTAgTCT	22081
188LeuProIleLysAlaMetL	194
22080	TTCCTTtCAGCCcAGTtCTCTCGAGCAGCTGTCCCATtAAAGGcGARtGG	22031
194	euaRGserLeuProIleTriPserIlEcysLeuGIyCysPheSerHsgln	210
:::		
22030	tCACATtGCCtTACCACtTtGGGCCATtTtCTGGStTTTTttTAGCcATtTtC	21981
211	TriLeuValSerThrMeLValValTyrlleProThrTyrlleSerSerVa	227
	:::	
21980	TGtGTtGGtCAGCCATtCATtCTtAACATtACCTtACtACAGtATtATtACtAGtACTt	21931
227	IYrHSVAIAshIIeArGASP.....	234
:::		
21930	GCTCCATGTTAACATCAGAGATGTGAgtTtACTTCCtATACTtCPtACGAA	21881
234	234
21880	AATGATAAtgTAATAAGAAGAAACAGTtCTGTGTtACCTtTtACATtCT	21831
234	234
21830	GGCTTtACATtAAcCATtAAtTTAAcCTtCACAAtGACtTGAGAGAG	21781
234	234
21780	CATtGTtAAtATtCCtTtTtCACAGATGTGMAACAGAGACtTtAGAGct	21731
234	234
21730	GAGATtACTtGGCCcCAGtTGcACAAtACTtAAGtAGtAGActGCTGcAG	21681
234	234
21680	CATtCATtATtCTtAAcCAtTAtGtATtACTtACACAGcAGcTGAgtTtCAA	21631
234	234
21630	AcCTtCTtTtTAGAAAtAAAtATtTtCTGTGGcCAGcAGcATGtGTcATGcCTtG	21581
234	234
21580	TAATtCCACACACTtTtGGGAGcCCGAGcAGcAGcAGcATcATcAGtGTcAGGA	21531
234	234

21530	TGCAAGACACGCGCTGACCAATATGGTTTACTAAATATCATCTACTACAAAAA	21461
234	234
21480	TACAAAAATTAGCCAGGTGTGTGGCAGGACCGACTGTAATCCAGCATTTC	21431
234	234
21430	AGGAGCTGAGACAGGAGATCGTTGAACCGAGAGGTGGAGTTGCAT	21381
234	234
21380	TGAGCCAGATCATGTGCACATGCACATCCAGCCTGGCGACAGACTAAGACT	21331
234	234
21330	CCGTTTCAAAAAACAAAAACCAAGAAATTAAATTGTTTATCTTGAGAG	21281
234	234
21280	CCGAGTGTAGTCAGCTGCTGSCCCTTATCTGAGACAGTGTCTTTTA	21231
234	234
21230	GTTGTAAAAAGATGTCTATTTTCCCCAAACAACCAAGTATCATGGG	21181
234	234
21180	GGTAACTTAATGGCTGCTGCTGTACTACACAATTTTGGTCTAACTGA	21131
235Asn..GlyLeuLeus 239	
21130	TCTCTAATCACTCTGTGTAAACTTCCTCTTCAGAGTGGAGTTCTGT	21081
239	erAlaLeuProPheIleValAlaIrtPValIleGlyMetValGlyIleTyr	255
21080	CGTCCCTGCTTTTATTCGTGTCGCAAGCTGTACAATTTAGAGGTGAG	21031
256	LeuAlaSPheLeuLeuThrIlySlys..PheArgLeuIleThrValAr	271
21030	CTGGCAGATTCTCTTTTGTCCAGGAATCTTCAGATTGATCAGTGTGCG	20981
271	gIysIleAlaThrIleLeuGlySer..Leu.....	280
20980	AAAGCTTTTCACTCTTGTGAAGATAGAGCGTGTGGCCCATTTAAC	20931
280	280
20930	AATCCCTTTTGTGACATGTGTCAGAGGTTCCCTGACAGCATGTCTTC	20881
281Proser..SerAlaLeuIleValSerLeuProTyr	291
20880	ATTGCCAGGCGCTCCCTCCATCAATATGTGCTGGCCCTGCCCTTT	20831
292	LeuAsnSerGlyTyrIleThrAlaThrAlaLeuLeuThrLeuSerCysG1	308
20830	GTGGCCTCCAGTTAGCTGTAAACCATTAATTTTGCTATCTTAATTCCTGG	20781
308	YLeuSerThrLeuCysG1SerGlyIleTyrIleAsnValLeuAspIleA	325
20780	GACCACTAACCATATGTAGCTCAGGAGTTTATCATCAACACCTTAGATATCG	20731
325	laProArgTyrIleSerSerPheLeu 332	
20730	CCCCCAGGTAAAGCGTCTACCTG 20708	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 13:48:53 ; Search time 51.16 Seconds
(without alignments)
459.549 Million cell updates/sec

Title: US-09-391-958-1
Perfect score: 2076
Sequence: 1 MOVDETLIPRKVPFSLCSARY.....LIFGEADVQEWAKERKLRL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues
Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_63:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	964	46.4	467	2	A48916 sodium phosphate t
2	946	45.6	465	2	S68915 sodium-phosphate t
3	939	45.2	465	2	I39473 Na+-dependent phos
4	932	44.9	465	2	A56410 sodium/phosphate t
5	475	22.9	560	2	I59302 brain specific Na+
6	449	21.6	573	2	T23589 hypothetical prote
7	434	20.9	544	2	T24633 hypothetical prote
8	429	20.7	472	2	S28286 hypothetical prote
9	388	18.7	413	2	T01534 hypothetical prote
10	370	15.4	516	2	T24729 hypothetical prote
11	306.5	14.8	499	2	T15201 hypothetical prote
12	304.5	14.7	530	2	T29418 hypothetical prote
13	282.5	13.6	506	2	T29668 hypothetical prote
14	280	13.5	568	2	S44742 C02C2.4 protein -
15	279.5	13.5	420	2	S44900 ZK652.10 protein -
16	279.5	13.5	592	2	T25419 hypothetical prote
17	278.5	13.4	450	2	A65061 hypothetical prote
18	259	12.5	478	2	T29174 hypothetical prote
19	258	12.4	444	2	C65102 hypothetical 49.0
20	254.5	12.3	543	2	T32496 hypothetical prote
21	252.5	12.2	455	1	H69752 probable glucarat
22	250	12.0	466	2	S40767 hypothetical prote
23	247.5	11.9	516	2	T27092 hypothetical prote
24	247	11.9	445	1	D65171 hypothetical 48.8
25	247	11.9	537	2	T20746 hypothetical prote
26	244	11.8	455	2	T34366 hypothetical prote
27	244	11.8	473	2	T31717 hypothetical prote
28	243	11.7	493	2	T19383 hypothetical prote
29	229	11.0	659	2	T33557 hypothetical prote
30	227.5	11.0	462	2	T34365 hypothetical prote

ALIGNMENTS

RESULT 1	31	224.5	10.8	453	1	F65250	hypothetical 49.4
A48916	32	224	10.8	435	1	S65882	transport protein
sodium phosphate transport protein 1, renal - human	33	219.5	10.6	419	2	T19260	hypothetical prote
C:Species: Homo sapiens (man)	34	218	10.5	479	2	T29720	hypothetical prote
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999	35	216	10.4	478	2	T33942	hypothetical prote
C:Accession: A48916	36	211.5	10.2	485	2	T24115	hypothetical prote
R:Chong, S.S.; Kristjansson, K.; Zoghbi, H.Y.; Hughes, M.R.	37	206.5	9.9	493	2	T25357	hypothetical prote
Genomics 18, 355-359, 1993	38	205.5	9.9	421	2	T22869	hypothetical prote
A:Title: Molecular cloning of the cDNA encoding a human renal sodium phosphate transp	39	205.5	9.9	487	2	T23776	hypothetical prote
A:Reference number: A48916, M0ID:94117004	40	200.5	9.7	466	2	T32204	hypothetical prote
A:Accession: A48916	41	200	9.6	453	2	T30985	hypothetical prote
A:Status: preliminary	42	198	9.5	552	2	T23755	hypothetical prote
A:Molecule type: mRNA	43	197.5	9.5	472	1	B65098	hexonate transpo
A:Residues: 1-467 <CHO>	44	190.5	9.2	422	1	A69853	hexonate transpo
A:Cross-references: GB:X71355; NID:q450531; PIDN:CA50490.1; PID:q450532	45	185	8.9	467	2	T16319	hypothetical prote
A:Gene: GDB:SLC17A1; NP1							
A:Cross-references: GDB:141889; OMIM:182308							
A:Map position: 6p23-6p21.3							
Query Match	46.4%	Score 964	DB 2	Length 467			
Best Local Similarity	41.3%	Pred. No. 1.8e-69					
Matches 193; Conservative 68; Mismatches 140; Indels 66; Gaps 3;							
QY 1 MOVDETLIPRKVPFSLCSARYGIALVLFHCNFTTAQNVNITMVAWVNSTSPQSOLNDS 60							
DB 1 MOKDNRLPKKVPFCFSFRYGLSFVHCNVITTAQACLNLMVYVNSTDPDHGLPNTS 60							
QY 61 SEYLV-----PVDSFGLSAKPKSL----- 80							
DB 61 TKRLDNINKNPMYNSPDIIGILSTSYGVIIIOVPVGFSGYKMKIGALCLSSV 120							
QY 80 -----PAKSSILGOFIMEMGSPQESRSCIALSGML 115							
DB 121 LSLILPPAGIGVAVVYVCAVGAAGIVATQFEIYVWAPPLEGRKLTSMSTSGFL 180							
QY 116 GCFETAILIGFISETLGWPFVFIYFGVGVCVCLLMFVYIDDPVSYPWISTEKEKTIIS 175							
DB 181 GPFIVLVYLVGVIESLQGWPMVFYIFGACGACVCLLMVFLYDDPKDKHPCISIEKEKTIIS 240							
QY 176 SLKQOVGSSKQPLPIKAMRLSPDIWISICLQCFSHQWLVTMVVYIPTYISSVYHVNIRDN 235							
DB 241 SLVQOVSSSSKQSLPIKAILKSLPVMAISISFFEFWSHNINMTLYTPWFINSMLHVNKEN 300							
QY 236 GLTSALPEFIYAVYIGWVGVLADFLTKK--FRLLTVKIKITIGSLPSSALIVSLPLNS 294							
DB 301 GFLSSLPYLFAMVIGNAGLSDPFLRNLISVIAVAKLFTLAAGFLPALFEGVCLPLSS 360							
QY 295 GYTATATALLTSCGLSTLCOSGIYINVLDIAPRYSFSLMGASRGFSSIAVYPTVSGFL 354							
DB 361 TFPYSIVYFLILAGATGSPFCGIVGFINGLDIAPRYPFGIKKCSLTLMIGGLINSTLTGLI 420							

Db	121	IFIDQAAVGETWIIIVCRVVGITQGTVTTAQHEIHWKMAPRLEGRUTISMSLSGFLGP	180
QY	118	FTALLIGGFISEFTLGNPPVPIIFEGVGCVCLLMFPVYIDDPVSPWISTEKEXIISL	177
Db	181	FVTLVGLGIIICESGMPWPVEYIFGACGVCCLLMFPVYIDDPKHPCCVSLHEKEYITSL	240
QY	178	KQVGGSKQDPLPIKAMLRSLPIWISICLGFESHQWLVTSMVYIIFTYIISVYHNIRDGL	237
Db	241	IQGSSFTROSLPIKAMIKSLPLMAISFCCFAYLWTSRLIYPTPLINSMHVDIREGL	300
QY	238	LSALPEIVAMVIGWVGGLADELLTKKFRULT-YRKIATILGSLPSSALIYSLEPLNGY	296
Db	301	LSSLPYLPFAWICGVYAGTADPLMSRNMLSTAIKRKLETAIGLLPIYFSCCLLYLSSGF	360
QY	297	ITATALLITLSCGLSTLCOSGIYINVLDIAPRYSSEFLMGASRGFSSIAVPIYPTGFFLS	356
Db	361	YSTTFPLLIANAASSFCIGALINALDLAPRYVEYIKQVTLTLCMTGGMSTSVAGPLFS	420
QY	357	ODPEFGMRNVEFLFAVNLGLLFYLTGEGADNOEMAKERTLSL	401
Db	421	ODPESSWFKITFLMSIINWISVITFLPAKAEIDMAEKQHTRL	465

RESULT 5
I59302
brain specific Na⁺-dependent inorganic phosphate cotransporter - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence-revision 02
C:Accession: I59302
C:NI, B.: Rostreck, P. R.; Nadl, N. S.; Paul, S. M.
Proc. Natl. Acad. Sci. U.S.A. 91, 5607-5611, 1994
A:Title: Cloning and expression of a cDNA encoding a brain-specific Na(+)-dependent inorganic phosphate cotransporter
A:Reference number: I59302; MUID:94261635
A:Accession: I59302
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-560 <RES>
A:Cross-References: EMBL:U07609; NID:9507414; PIDN:AA19646.1; PID:9507415

```

Query Match 22.9% Score 475; DB 2; Length 560;
Best Local Similarity 28.0% Pred. No. 2.2e-30;
Matches 111; Conservative 66; Mismatches 158; Indels 62; Gaps 6;

QY 12 VPSLCARVGIATLVLFHCNFTTIAQNVIMNITMVAWVNSTSPOSOLANSDSEYLPVDSFG 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 IPSARVHYGCVI-----FVRIQLGIVEGYTPAC----- 191

QY 72 LSKAPKSLPAKSSILGCGFAIWERKNGPPOERSRLCSIALSCMLGCFPAIIIGCFISTEL 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 -----HGIMSKWAPLERSRIATTFQCSYAGAVVAMPDLAVQYS 232

QY 132 GMPFVFYIFGGVGCVCCLMFVYIYDDVSPYMIISTSEKEITISLKKQVSS----- 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 GMSVFYIYFGSGFIFWYIFMLIVSYESPALRPSISEEKRII-----EPAIESAKLMPV 288

QY 185 -KQPLIKAMLRSLPIMSICLGCFSHOWLVSTMVYIIPYIISVYHVNIRONGLSALPF 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 TKFPMTPMRFFETSMFVYAIIVANFCRSWTFYLLISQPAYEEVGFISKVLGVSALPH 348

QY 244 IVAWIGWVGVVLDELTKRKRLLT-VRKATATIGSLPSNALLYSLPYLMSGYTATAL 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 LVMTFIVYIGQIADFLSRHIMSTTNVTKLNCGGFGMEATFLLLVAYSISKGV-AISF 407

QY 303 LTLSCGLSTLCOSGIYIVNDIAPRYSFLMGASRGFSFAPVIYPTVSGFLISODPREG 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 LVLANGFSGFAISGFNVNHLDIAPRYASITILMGISNGVETLSGMCPILVGMTAKTKTREE 467

QY 363 WRNVEFLFAVNLGLFYLILFEGADYQEWAKERLIT 399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 WQYVFLIASLVHYGVIFYGVFASGEKQPMAPDEEMS 504

```

RESULT 6
T23589
hypochemical protein K10G9.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T23589; T24636
R: Mortimore, B.
submitted to the EMBL Data Library, August 1994
A: Reference number: Z19765
A: Accession: T23589
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-573 <WIL>
A: Cross-references: EMBL:Z36282; PIDN:CAA85289.1; GSPDB:GN00021; CESP:K10G9.1
A: Experimental source: clone K10G9
R: Buck, D.
submitted to the EMBL Data Library, February 1995
A: Reference number: Z19915
A: Accession: T24636
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-573 <WIL>
A: Cross-references: EMBL:Z48055; PIDN:CAA86135.1; GSPDB:GN00021; CESP:K10G9.1
A: Experimental source: clone T07A5
C: Genetics:
A: Gene: CESP:K10G9.1
A: Map position: 3
A: Introns: 38/3; 87/3; 224/1; 272/3; 322/3; 444/3; 476/3; 496/1; 517/2

Query Match	21.6%	Score 449	DB 2	Length 573
Best Local Similarity	26.7%	Pred. No. 2.6e-28		
Matches 125; Conservative	84;	Mismatches 165;	Indels 94;	Gaps 14

```

Oy      1  MOVDETLIPRKPSLCSAYGIALYLHF-----CHETIAQVININIVMAVNST  51
Db      35  IQEDKKKFLPKRY-----RMQIALLAHREFALSPGIRSNF-GVAKRMNIN-----NFT  80
Oy      52  SPQSLAND-----SSEVLPVDS--FGGLSK-----APKSLPAKSSILGQFA----  92
Db      81  DAYGEVHEKEFPWGTGEVGMESSEFFYGAAAGQIPAGVIAAKFAPNKLPLMGLIFASLLN  140
Oy      92  -----
Db     141  IYTAICLNHPPTDIFVMVIVQMOGLALGVCYPRAMHGVMKWYAPLERSKLATTTFTGAS  200
Oy     115  IGCFFAILIGRISLSETLGMFVFYIFGCGVCVCCILMFVYIYDDPSPYMWISSEKEYII  174
Db     201  VGVWVGIPASALVALSHFSSTPFPYFGALGIWMSILMIYVSGTSPTETHGISADEKKYI-  260
Oy     175  SLSKQOVGS-----SKOPLPIKAMLRSLPIWSICLGCPSHOMLVSTWVVIPTYISSVY  228
Db     260  ---TEKVSVAVKNNMTLTLLTPMRDMMTSTAAVVAIIICSCRCRSHSPFLILGNOLITWKDY  316
Oy     229  HYNIRDNGLLSALPEITVAVNIGMVGGYIADFLTK-KFRLITVRKIATIIIGSLSPSALIV  287
Db     317  HDIKNSGILAIIPQLGMCIVTLTSGQLSDLYRSSGKMSTEAVRKSVNPFGEVTEAVMIG  376
Oy     288  SLPIYNSGTYITFALITLISGSLTQOSGITYINVLDIAPRYSSFLMGASRGFSSIAPIVY  347
Db     377  CLAFPRDPIAVIATFLI-IAGSGAGAVLSGFENVNHHEDIARHAPIILMIGANGAIGAV-G  434
Oy     348  PTVSGFLLSODEPFGMRNVFLLPFAVNILGLFVYIFEGADVQEMAKE  395
Db     435  GIYVNSLTYQNPD-GMOWVFLIAMSIDIFGIITFLIFAKGDVLPMAKE  481

RESULT      7
T24633
hypotheitcal protein T07A5.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24633

```

Query Match 20.7%; Score 429; DB 2; Length 472;
Best Local Similarity 27.9%; Pred. No. 8.3e-27;
Matches 109; Conservative 83; Mismatches 162; Indels 36; Gaps 12;

```

Db      279 SLTFNIFCEQVLKFNLTESGLLCPWILMAVFANIGMIADTLVSRG----- 327
Oy      277 LGLSSALIVSLPLNLSGYITATATLTLSCSLSTLCOSGIYINVDIAPRYSFLMGAS 336
Db      327 -----LSTTNGSDAFSOSGLTSMNODITPRYAGVLLGLIS 360
Oy      337 RGFSSIAPIVPTVSGFLLSDPFGMRVFFLLFAVNLGLLEFYLIEGAD 388
Db      361 NTAGVLACVFGTAATGCTILGRG---SMDVFRVAVALYLIGLWNLFRATGE 409

```

RESULT 10
T24729
hypotheical protein T09B9.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24729
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: 219929
A:Accession: T24729
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-516 <MIL>
A:Cross-references: EMBL:247070; PIDN:CAAB7340.1; GSPDB:GN00028; CESP:T09B9.2
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: CESP:T09B9.2
A:Map position: 10
A:introns: 85/1; 143/2; 184/1; 241/3; 344/1; 391/1; 436/2; 470/3

Query Match 15.4%; Score 320; DB 2; Length 516;
Best Local Similarity 25.4%; Pred. No. 4,5e-18;
Matches 105; Conservative 78; Mismatches 177; Indels 54; Gaps 13;

```

Oy      29 CN-----FTTAQNYI---MNTIVAMVNSTSPQSOLD-----SSRFLPV 66
Db      71 CNHPRQFAMSSIQGLITISGNGFSLPMVITGMQADRLKMTIYAAMAFTIVSNAVLPT 130
Oy      67 D-----SFG-----GLSKAPKSLPAKSSILGQFAIERNRGPPOERSRLCSIALSGML 114
Db      131 SAGAFALVFELRYLTFEGDALLS-PASSSL-----TRNFPFRERSALGIYTSGRQ 182
Oy      115 LCCFTAILIGGT-----SETL-GMPFVYITFGVGCVCCLLMFYVYIYDPSYPISTIS 168
Db      183 IGTLLIPLPGWLCGSDGSKFLGMPALFYLLSSVAAALVIMVVFSAADKPSKHLICISHN 242
Oy      169 EKEYIISLKOO-VG--SSKQPLPIKAMRLSLPIWSICLGCFSHOMLVSTWVYIPTIYS 225
Db      243 EEAIVNKRKIEENIGKRRNRKNTPKALFTSKOVAAVALVCHERPLVIMQLFKPFS 302
Oy      226 SVYHAVNIRNDGLLSALPFIYAVVIGVGGYLAFLTLTKKF-RLITVRKIATTLISLPSSA 284
Db      303 DVLGSLNFTVNGVLSALPMAILFLSKCLASLASLYLANGYLTKQSCKIFNFIASGLGI 362
Oy      285 LIVSLPYLNS--GYITATALLTSCGLSTLCOSGIYINVDIAPRYSFLMGASRGFSSI 342
Db      363 CIAAATPLMSNLOHAIWAILIICLIANAFGLHRPGVITATVQALAPFSGIITGIAPAVASC 422
Oy      343 APVIPTVSGFLLSDPFGMRVFFLLFAVNLGLLEFYLIEGADVQMAKER 396
Db      423 PSIFPKLLISQILTRGSKHEMTIVEISAFVALPIIFTLWGSABRTMASNR 476

```

RESULT 11
T15201
hypotheical protein F12B6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15201
R:Pauley, A.; Mag91, L.

submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid F12B6.
A:Reference number: 218307
A:Accession: T15201
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-499 <PAU>
A:Cross-references: EMBL:AF003138; NID:92088708; PID:92088710; PIDN:AAB54154.1; GSPDB
A:Experimental source: strain Bristol N2; clone F12B6
C:Genetics:
A:Gene: CESP:F12B6.2
A:Map position: 1
A:introns: 23/1; 62/2; 75/2; 117/2; 163/1; 277/3; 319/3; 363/1; 413/3; 443/3

Query Match 14.8%; Score 306.5; DB 2; Length 499;
Best Local Similarity 24.5%; Pred. No. 5,2e-17;
Matches 80; Conservative 72; Mismatches 134; Indels 41; Gaps 7;

```

Oy      79 LPKSSILGQFAIERNRGPPOERSRLCSIALSGMLGCFATLIGGFISLTL----- 132
Db      166 IPGGSVLIS-----WPLSEKSTAMALFTTGNOIG-----IAMSFLAKLQQLHFE 214
Oy      132 GMPFVYITFGVGCVCCLLMFYVYIYDPSYPISTSEKEYIISLKOOVSSKQPLPIK 191
Db      215 GMPLVYITVGLGAVPLVIMHVRMLADKPRESKYITATLTYI-----KGGQRNRRAE 267
Oy      192 AMRSLP-----TMSICLGCFSHOMLVSTWVYIPFYISSVYHVNIRNDGLLSALP 242
Db      268 TIVRATPYMKIILNGCVACICAFQSFQSFVLAVALYTLPRKYNQIAFKMNLTHNGIWSSLP 327
Oy      243 FIYAVVIGVGGYLAFLTLTKKRLITVRKIATTLISLPSNALIVSLPLNLSGYITATL 302
Db      328 FFIQMTKLLPAILADKVKQRKNATVTKVSNALISFASALFTVIAA---GPFDSABL 384
Oy      303 LFLSCGLSTLCOS---GIYINVDIAPRYSFLMGASRGFSSIAPIVPTVSGFLLSOD 358
Db      385 VQLSIVSMAASAVPYPGYNTSIVYVAPQFTAFISSYQALYNQIASTAPIYIGRTISHG 444
Oy      359 PERGMRNFFLLFAVNLGLLEFYLIFG 385
Db      445 TIYEMKCAFYSLAGVLAATGLIFQIFG 471

```

RESULT 12
T29418
hypotheical protein F21F8.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Wilson, R.; Favello, A.; Le, T.T.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid F21F8.
A:Reference number: 220618
A:Accession: T29418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-530 <MIL>
A:Cross-references: EMBL:097000; PIDN:AA04797.1; GSPDB:GN00023; CESP:F21F8.11
A:Experimental source: strain Bristol N2; clone F21F8
C:Genetics:
A:Gene: CESP:F21F8.11
A:Map position: 5
A:introns: 16/1; 64/2; 112/3; 344/3; 417/1; 452/1

Query Match 14.7%; Score 304.5; DB 2; Length 530;
Best Local Similarity 22.9%; Pred. No. 8e-17;
Matches 110; Conservative 69; Mismatches 186; Indels 115; Gaps 15;

```

Oy      15 LGSARYGIALVLR-FCNFTTAQNYIIN-----ITWVAMVNSTSPQSOLNDS 60
Db      38 MCOMFAHIGLSLSCMCNSTAVA---LMTNNTATLVGTESTILMSILEKNTSESQELGGG 94

```

QY 61 S-----EVLPSVDSGG----- 72
| : : : : :
Db 95 DDGESCRLBEKVIKDGRTIMSVWOGYIVSAFLGGLFISIPAGVLDVDFSAKHLLS 154
QY 72 -----LSKAPKSLPAKSSILG--GQFA-----IMERGPPQERSRLC 106
| : : : : :
Db 155 VAILMLTIALSLMPLYSIYIEKGAFAGRFVWGISEMLPISINSMTWKMPINKSKSLA 214
QY 107 SIALSG-MLLGCFTAILIGGFISETLGGPFVYIFGCGVCCCLMFVYIYDDPVSYIWI 165
| : : : : :
Db 215 SVTAGNGLSGMFGNIIAELCASSFGSSIFYSASLEIGIMLVMHLLTVRNSPHNTIWI 274
QY 166 STEKEVIISLKOQVSSS-KOPLPIKAMLSLPIWISICLGFSGHQLVSTWVYIPIYI 224
| : : : : :
Db 275 HKRELDIYANNIPRHPSSVYKKTIPWMLTSKYVMSLMPMSVGMGMALIFVYIPIYF 334
QY 225 SSYYHVNIIRDNGLLSALPFIYAWVIGVGYLADFLTKRF---RLITVRKIATILGSL 280
| : : : : :
Db 335 KDVLMLDVQSNFGYSAIPHISNLIAKLIWGLMDKMRKILLSPSATVYKLSQFASMG-- 393
QY 281 PSSALIVSLPLNSGYIAT---ALITLSC-----GLSTLQSGIYINVLIDAPRYSFL 332
| : : : : :
Db 393 -ISVSCFELRYMN---CATPFYALVLLSVSAFEGLSI---SGEYTSLSIAPSHIGTL 444
QY 333 MGASRGFSIAPVIYVYSGFLLSODPEFGMRNPFELFAVNLGLLYLIFEGEADVOEW 392
| : : : : :
Db 445 TSLATVIGFVGMRPFLPMISTYKTYGTAEKMGHLLIYVFASATGCIIFLFGSGDVQNW 504

RESULT 13

229968
hypothetical protein ZK682.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29968
R:Du, Z.; Le, T.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid ZK682.
A:Reference number: 220714
A:Accession: T29968
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-506 <DU2>
A:Cross-references: EMBL:U01110; PIDN:AAA82415.1; CESP:ZK682.2
A:Gene: CESP:ZK682.2
A:introns: 37/3; 82/3; 163/3; 314/3; 359/1; 469/1

Query Match 13.6%; Score 282.5; DB 2; Length 506;
Best Local Similarity 24.6%; Pred. No. 4.3e-15;
Matches 82; Conservative 70; Mismatches 140; Indels 41; Gaps 11;

QY 85 ILG-GQFAIW-----ERMGPPQERSRLCSIALSGMLLGFETAILIGFISETL----- 132
| : : : : :
Db 158 VMFGQGVLMPCMLVLIAQWFPVNEKSTALAIATGNQL---SVYIAMFATAEICQLPW 213
QY 132 GMPFVYIFGCGVCCCLMFVYIYDDPV-STPWISTEKEYIISLKOQVSSSKOPLPI 190
| : : : : :
Db 214 GMPAFHYAVACGIWCIWYMMVYDSCPHADKLSRDELHIIT---ERVILRPQHNMW 270
QY 191 KAMRLSLPWSICLGFSGHQLVSTWVYIPIYISVYHVNIIRDNGLLSALPFIYAWVIG 250
| : : : : :
Db 271 MALMKSPPVWMLAASSFPHNVYTGTTIYLLPYKTYVNMSTLSNGLMSALPEVQLLSK 330
QY 251 MWGGLADFLTKRFLITVRKIATILGSLPSSALIVLSLPY-----NSGYITATALLT 304
| : : : : :
Db 331 VFYAGMAE--SARKRDWDINKRITFCNSASFGIATCGLLCDDCCORG-----AAIF 383
QY 305 LSGGLSTLCQSGIYI-----NVLDIAPRYSFLMGASRGFSIAPVIYVPSGFLSDP 359
| : : : : :
Db 384 LIC--LAMCFVSGYIPGYNTSAVTIAPGQTAIAAFSFRWQIASSVAPYHIGAYTKOGT 441

QY 360 EFGMRNPFELFAVNLGLLYLIFEGEADVOEW 392
| : : : : :
Db 442 ADEWKIVFAVIAICVGVGIFQCGGTASLDOW 474

RESULT 14

544742
C02C2.4 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C:Accession: S44742
R:Wilson, R.
submitted to the EMBL Data Library, September 1993
A:Description: Sequence of the C. elegans cosmid C02C2.
A:Reference number: S44737
A:Accession: S44742
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <WIL>
A:Cross-references: EMBL:L23649; NID:9388566; PID:9388572
C:Genetics:
A:introns: 39/1; 55/3; 103/3; 128/3; 206/3; 266/1; 310/3; 361/3; 406/1; 436/2; 487/3

Query Match 13.5%; Score 280; DB 2; Length 568;
Best Local Similarity 22.8%; Pred. No. 7.8e-15;
Matches 92; Conservative 68; Mismatches 191; Indels 52; Gaps 10;

QY 5 ETLIPKVPYSLCSARYGIALV-----HFCNFTTIAONVIMNTVMVNSTSPQSOLN 58
| : : : : :
Db 161 ERFSPRHVLQISVALYILVYITPELATHFGYFVFLARIGM----- 203
QY 59 DSSEVLPVDSFGGLSKAPKSLPAKSSILGQPAIEMRGPPQERSRLCST-ALSGMLIGC 117
| : : : : :
Db 203 -----GLQEG-FVEPTNNAIIG-----NMFPSEKSTALSIPTLGNQISA 242
QY 118 FTAILIGFISPTLGGMPVYIFGCGVCCCLMFVYIYDDPVSYPMWISTEKEYIISL 177
| : : : : :
Db 243 AGSPWAAVACASDLQWPTTFPAGIFATGWSITLMEFTTASHPAKYKMTKKEKYLNAV 302
QY 178 KOQVSSK--QPLPIKAMLSLPIWISICLGC-FSHQMLVSTWVYIPIYISSVYHVNIIRD 234
| : : : : :
Db 303 VKKVKSEKTRSIYPSKILTS-PAFLGQLCHFFVNLPMFLFOYILPSEYFVHLGVIA 361
QY 235 NGLLSALPFIYAWVIGVGYLADFLTKRRLIT-VAKIATILGSLPSSALIVSLPLN 293
| : : : : :
Db 362 NGFTALPNIIFNMILFKVYWGIGIDLKENKILLSNRKAVSHGVASFGSSFSLLILAFV 421
QY 294 SGYITATALLTSCGLSTL--CQSGIYINVLIDAPRYSFLMGASRGFSIAPVIYPTVS 351
| : : : : :
Db 422 DCSNPTTGLIFCLMYTSMGTFVSGFYTSLSIAPQYIATYSALSMFPAAMIGRLTTPAVM 481
QY 352 GFLLSQDEPFGRNVEFLLFAVNLGLLYLIFEGEADVOEMAK 394
| : : : : :
Db 482 SMFRKDTAAEWMONIFIGCSLAHIFSGSIFLLFGSGELQDMAK 524

RESULT 15

544900
ZK652.10 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 09-Sep-1997
C:Accession: S44900
R:Du, Z.
submitted to the EMBL Data Library, May 1993
A:Description: Sequence of the C. elegans cosmid ZK652.
A:Reference number: S44613
A:Accession: S44900
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <DU2>
A:Cross-references: EMBL:L14429; NID:9289765; PID:9289766

THIS PAGE BLANK (USPTO)

OY	61	SEVLPPDSSGGJLSKAKSPISPAKSSIIGGOFATIERBNGPOERSRCLSIASGLMGCFTA	120
Db	61	SEVLPPDSSGGJLSKAKSPISPAKSSIIGGOFATIERBNGPOERSRCLSIASGLMGCFTA	120
OY	121	ILIGGFISFTLGMPEVFEYIFGGVGCCLLMFVIVYIDDPVSYPWISTSEKEYITISSLKOO	180
Db	121	ILIGGFISFTLGMPEVFEYIFGGVGCCLLMFVIVYIDDPVSYPWISTSEKEYITISSLKOO	180
OY	181	VSSSKOPFLIKAMLRSLPMSICLGGEFHSOMLVSTWVVYIPMTISSVYHNRDNLLSA	240
Db	181	VSSSKOPFLIKAMLRSLPMSICLGGEFHSOMLVSTWVVYIPMTISSVYHNRDNLLSA	240
OY	241	LPIFIYAVMIGNGVYIADBLTLTKRFFLIYVRKATIILGSLPSALLVSPYLNSGITAT	300
Db	241	LPIFIYAVMIGNGVYIADBLTLTKRFFLIYVRKATIILGSLPSALLVSPYLNSGITAT	300
OY	301	ALILTSCGSTLCQSGITINVDIAPRYSFTLMGASRGFSSTAPVIPTPVSGFLSQDBE	360
Db	301	ALILTSCGSTLCQSGITINVDIAPRYSFTLMGASRGFSSTAPVIPTPVSGFLSQDBE	360
OY	361	FGMRNVFLLFAVNILLGLFYILFGADVQENAKKKRLRL	401
Db	361	FGMRNVFLLFAVNILLGLFYILFGADVQENAKKKRLRL	401
RESULT	2		
ID	NPT1_HUMAN	STANDARD:	PRT; 465 AA.
AC	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	15-OUL-1998	(Rel. 36, Last annotation update)	
DE	RENAL SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 1 (SODIUM/PHOSPHATE COTRANSPORTER 1) (NA+)/PI COTRANSPORTER 1) (RENAL SODIUM-PHOSPHATE TRANSPORT PROTEIN 1) (RENAL NA+-DEPENDENT PHOSPHATE COTRANSPORTER 1) (NA/Pi-4).		
DE	SLOC17A1 OR NPT1.		
CN	Homo sapiens (Human).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Primates; Catarrhini; Homidae; Homo.		
NC	[1]		
RN	SEQUENCE FROM N.A.		
RP	TISSUE=KIDNEY.		
RC	MEDLINE: 94117004.		
RA	Chong S.S., Kristjansson K., Zoghbi H.Y., Hughes M.R.;		
RT	"Molecular cloning of the cDNA encoding a human renal sodium phosphate transport protein and its assignment to chromosome 6p21.3-p23."		
RT	Genomics 18:355-359(1993).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	TISSUE=KIDNEY CORTEX;		
RC	MEDLINE: 95126933.		
RA	Miyamoto K.-I., Tatsumi S., Sonoda T., Yamamoto H., Minami H.,		
RT	Taketani Y., Takeda E.;		
RT	"Cloning and functional expression of a Na(+)-dependent phosphate co-		
RT	transporter from human kidney: cDNA cloning and functional		
RT	expression." J. Biochem. J. 305:81-85(1995).		
CC	- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.		
CC	MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA		
CC	Na+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.		
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.		
CC	- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY CORTEX, LIVER AND BRAIN		
CC	BUT NOT IN OTHER TISSUES.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		

```

CC      EMBL: X71355; CAAS0490.1; ALT-INT.
DR      EMBL: D28532; BAA05888.1; -.
DR      MIM: 182308; -.
DR      PFM: PF00083; sugar_tr; 1.
KW      Transport; Transmembrane; Glycoprotein; Symport; sodium transport.
FT      TRANSMEM 79      99      POTENTIAL.
FT      TRANSMEM 117     137      POTENTIAL.
FT      TRANSMEM 176     196      POTENTIAL.
FT      TRANSMEM 198     218      POTENTIAL.
FT      TRANSMEM 255     275      POTENTIAL.
FT      TRANSMEM 299     319      POTENTIAL.
FT      TRANSMEM 337     357      POTENTIAL.
FT      TRANSMEM 363     383      POTENTIAL.
FT      TRANSMEM 399     419      POTENTIAL.
FT      TRANSMEM 431     451      POTENTIAL.
FT      CARBOHYD 39      39      POTENTIAL.
FT      CARBOHYD 47      47      POTENTIAL.
FT      CARBOHYD 56      56      POTENTIAL.
FT      CONFLICT 35      36      RA -> LM (IN REF. 2).
FT      CONFLICT 77      77      I -> V (IN REF. 2).
FT      CONFLICT 207     207      G -> C (IN REF. 2).
FT      CONFLICT 231     231      S -> G (IN REF. 2).
SQ      SEQUENCE 465 AA; 50884 MW; 687A19F245C4ED16 CRC64;

Query Match          46.0%; Score 954; DB 1; Length 465;
Best Local Similarity 41.1%; Pred. Mism. 6,5e-65;
Matches 191; Conservative 68; No matches 140; Indels 66; Gaps 3

QY      3 VDETIIPKRVPLCSARVGIALVHFCNFTTTAQNVINMTVMVAVNSTSPQSOLDNSE 62
DB      1 MKNRLPKKVPKVCFSRYGLSFLVHCCNVITTAQRACNLMLVMVAVNSTDPGLPNTSRK 60
QY      63 VL-----PYDSFGGLSKAPKSL----- 80
DB      61 KLLDNIKPMYMWSPDIOGILSSTSYGVIIIQVPGYSGSYSTRKMTGFALCLSYLS 120
QY      80 -----PAKSSILGGQFAIMERGMPQOERSRCLSIASLGMLGC 117
DB      121 LLIPPAAGCGVAVVYVCRAVGAAGCIYATAPETIYKAPPLERGLTSMSTSGILLCP 180
QY      118 FPAALIGGFISTLGMPEFYVIFGGVGCYCLLMFVYIDDPVSYPWISTSEKEYITSSL 177
DB      181 FIVLLVTVGTGICSLQMPMYEYIFGACGCAVCLLMFVLYDDPKDHCISISEKEYITSSL 240
QY      178 KOQVSSKQPIPKMLRSLPIMSICLGGFSQHWLVSTWVVIIPYISSVYHVNIRNCL 237
DB      241 VOQVSSRSQSLPIKAILSLSPWALISIGSFTEFWSHNITLTTPMETNSMLHVNKENGF 300
QY      238 ISALPIYAVWGVAGVGIADFLTLTK- FRLTIVKRIATILLSLSSALIVSLPYLNSGY 296
DB      301 LSLPLYLPAWICGNLAGQLSDFFLTRNLISLVAVKRLFTPAAGFLILPAITGCVCLPYLSSTF 360
QY      297 ITATLLHLISGLSTLCOGSIITVLDIAPRYSFSLMGASRQSSIAPIYVTVSGFLIS 356
DB      361 YSIVFLILAGATGSRCLGAGVINGINDIAPRFGFIRKACSTLTGMIGGLIATSLTGLLIK 420
QY      357 GDFEGGRNVFPLFAVNLGLFLFLICEAVYQEMAKERKLTFL 401
DB      421 QDPESAMKRTFLTMAAINVTGLIFLIYATAIQDMAKEKHQTRL 465

RESULT      3
NPTL_RAT    STANDARD;      PRT; 465 AA.
AC      062795;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUL-1998 (Rel. 35, Last annotation update)
DE      RENA1 SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 1 (SODIUM/PHOSPHATE
DE      COTRANSPORTER 1) (NA(+)/PI COTRANSPORTER 1) (RENAL SODIUM-PHOSPHATE
DE      TRANSPORT PROTEIN 1) (RENAL NA+-DEPENDENT PHOSPHATE COTRANSPORTER 1)

```

```

GN SLIC17A1 OR NP11.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 97021433.
RA Li H., Xie Z.;
RT "Molecular cloning of two rat Na+/Pi cotransporters: evidence for
RT differential tissue expression of transcripts.";
RL Cell. Mol. Biol. Res. 41:451-460(1995).
CC -1- FUNCTION: IMPORTANT FOR THE RESORTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U28504; AAC52487.1; -
KM Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79
FT TRANSMEM 99
FT TRANSMEM 117
FT TRANSMEM 137
FT TRANSMEM 176
FT TRANSMEM 196
FT TRANSMEM 219
FT TRANSMEM 260
FT TRANSMEM 280
FT TRANSMEM 299
FT TRANSMEM 319
FT TRANSMEM 337
FT TRANSMEM 357
FT TRANSMEM 363
FT TRANSMEM 399
FT TRANSMEM 419
FT TRANSMEM 431
FT TRANSMEM 451
FT CARBOHYD 39
FT CARBOHYD 47
FT CARBOHYD 56
FT CARBOHYD 56
SQ SEQUENCE 465 AA; 51350 MW; 28BBB8DC5C0AC52 CRC64;

Query Match 46.0%; Score 954; DB 1; Length 465;
Best Local Similarity 41.5%; Pred. No. 6,5e-65;
Matches 193; Conservative 79; Mismatches 127; Indels 66; Gaps 6;

QY 3 VDETLIPRKVPSICSAARYGIALVHFCNFTTIAQNTVIMNTVMAVNSTSPQSOLNDS-S 61
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MEMRCPLKRVGFCSPFRYGALILHFCNIVYMAQVCLNTJTMAMVKKTEPRHLSKNSVA 60
QY 62 EVL-----PYDSF-----GGLSKA----- 76
   ||| |||
DB 61 EMDLVNKNPVHWSMDIOGLVLSVFLGAVNVIOVPVGLVSSVAPMEKIISSLSFLSVLS 120
QY 76 ----PKSLP-----AKSSILGGOFALMERMPQPERSRSLCISALSGMLGC 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 LILPPAONGALVIVCRVLQGIAGQAVSTGQHGIMVKNAPLERGLTMTLSGFWMP 180
QY 118 FTAILIGFISLTPMPFEVYIFGVGVCVCLMFVYIDDPVSPYPISTSEKEYITSSL 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 FIALIVSGFICDLGWMVMYIIFGIVGCVLSLFWFILLFPDPNNHNPMSSEKDYITSSL 240
QY 178 KOQVSGSKOPLPIKAMLRSLPIWSICIGCFSHQMLVSTWVVYIPYITSSVYHVNIRNGL 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 MOOVHSGROSLPIKAMKSLPLMAIILNSFAFIMSNMLVTPYPTFISTLHVAVRENGL 300
QY 238 LSLPLIVAMVIGMGVYLADELLTKK-FELIIVRKATITLSLPSALLVSPYLSNGY 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 LSLPLPLATIGVAGQMDPLLSKRTFSVAVRKLETTGLGFCPIVYVCLLYISYNF 360
QY 297 IFATALLTISCGISTCOSGIYINVLDIAPRYSSFLMGASRGSSIAPIVTPVSGFLS 366
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 361 YSTVIFILLANSTLSPFCQGLINADLADIRYGGFLKAVTALIGFCGLISSTLAGLILN 420
QY 357 QDEPFGKRVNFFLLFAVNLGLFLYIFGENDVQWAKKRLRL 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 QDEYAWHKNFFLMAGINVTCLAFYLLFAKGDIDQWAKKRTTL 465

RESULT 4
NP11_MOUSE
ID NP11_MOUSE STANDARD; PRT; 465 AA.
AC 061983;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RENAL SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 1 (SODIUM/PHOSPHATE
DE COTRANSPORTER 1) (NA+/-)/PI COTRANSPORTER 1) (RENAL SODIUM-PHOSPHATE
DE TRANSPORT PROTEIN 1) (RENAL NA+-DEPENDENT PHOSPHATE COTRANSPORTER 1).
GN SLIC17A1 OR NP11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE: 95335846.
RA Chong S.S., Korak C.A., Liu L., Kristjansson K., Dunn S.T.,
RA Boudreau J.E., Hughes M.R.;
RT "Cloning, genetic mapping, and expression analysis of a mouse renal
RL sodium-dependent phosphate cotransporter ";
CC Am. J. Physiol. 268:F1038-F1045(1995).
CC -1- FUNCTION: IMPORTANT FOR THE RESORTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77241; CAA54459.1; -
KM MGD; MGI:103209; SLIC17A1.
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79
FT TRANSMEM 99
FT TRANSMEM 117
FT TRANSMEM 137
FT TRANSMEM 176
FT TRANSMEM 196
FT TRANSMEM 219
FT TRANSMEM 260
FT TRANSMEM 280
FT TRANSMEM 304
FT TRANSMEM 324
FT TRANSMEM 337
FT TRANSMEM 356
FT TRANSMEM 363
FT TRANSMEM 383
FT TRANSMEM 399
FT TRANSMEM 419
FT TRANSMEM 429
FT TRANSMEM 449
FT CARBOHYD 39
FT CARBOHYD 47
FT CARBOHYD 56
FT CARBOHYD 56
SQ SEQUENCE 465 AA; 51589 MW; C67BE25A2C291EEF CRC64;

Query Match 45.6%; Score 946; DB 1; Length 465;
Best Local Similarity 40.2%; Pred. No. 2,6e-64;
Matches 187; Conservative 78; Mismatches 134; Indels 66; Gaps 3;

QY 3 VDETLIPRKVPSICSAARYGIALVHFCNFTTIAQNTVIMNTVMAVNSTSPQSOLNDS-- 61
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MEMRCPLKRVGFCSPFRYGALILHFCNIVYMAQVCLNTJTMAMVNNVTSPPHLSNESV 60
QY 61 -----SEVLPYDSFGGLSKAPSLP- 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 61 EMLDNVKNPVSNSPDIOGLILSSVEFGMVVQAPVGYLSGIYPMKRIIGSSILFLSLMS 120
OY 81 -----AKSIIIGGFAIWERNGPQPERRLCSIALSGMLGC 117
Db 121 LLIPPAOYGAALVYICRVLOGIACTVSTGHEIWMKAPLERGRILSMISLGVKMP 180
OY 118 FPAIIIGFISFTLGMPEVYIFEGVGCYCLLMFVYIDDDPSYMWISTSEKEYITSSL 177
Db 181 FIVLLVSGFICDILGMPWFYIFGIYGVLSLWFLFEDDDKDHVMSSSEKDIYITSL 240
OY 178 KOOVSSKQPLPIKAMLRSLPIWISICLGFSGHOMLVSTWVVIPIYISSYHVNRDGL 237
Db 241 MOOASSGRSLPIKAMLRSLPIWISICLGFSGHOMLVSTWVVIPIYISSYHVNRDGL 300
OY 238 ISALPFIYAWIGVYGLADFLTKK-PRILTVRKIATILSLSPSALIVSLPYNSGY 296
Db 301 LSLPYLAVIGIAGLQMSDFLRKISITVYKRLFTTSLSPCVITIMCLLYISYNF 360
OY 297 ITATALLTSLGSLTSCOSGIYINVLDAIPRYSFLMGASRGSSTIAPVYPTVSGFLS 356
Db 361 YSTVIFELIANSTLSFSGQLINLADIAPRYGFLKAVTALIGMGLISLTIAGLILN 420
OY 357 QDPFGMRNVFLLFAVNLGLFLYIFGEADVQEMAKERKILRL 401
Db 421 QDPESWPKIFLMSIINVISIVIFLIRAKAIDQMAKERTL 465

```

RESULT 5

NPT3_RABIT STANDARD; PRT; 465 AA.

```

AC 028722.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REMAL SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 1 (SODIUM/PHOSPHATE
DE COTRANSPORTER 1) (NA(+)/PI COTRANSPORTER 1) (RENAL SODIUM-POSPHATE
DE TRANSPORT PROTEIN 1) (RENAL NA+-DEPENDENT PHOSPHATE COTRANSPORTER 1)
DE (NMP-1).
GN SLIC17A1 OR NPT3.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY CORTEX;
RX MEDLINE; 92052140.
RA Werner A., Moore M.L., Mantei N., Biber J., Semenza G., Murer H.;
RT "Cloning and expression of cDNA for a Na/PI cotransport system of
RT kidney cortex."
RL Proc. Natl. Acad. Sci. U.S.A. 88:9608-9612(1991).
CC -1- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: KIDNEY CORTEX AND LIVER.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M76466; AAA31461.1;
CC PFAM: PF00083; sugar_tr; 1.

```

```

DR Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 109 129 POTENTIAL.
FT TRANSMEM 171 191 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
FT TRANSMEM 304 324 POTENTIAL.

```

```

FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 428 448 POTENTIAL.
FT CARBOHYD 47 47 POTENTIAL.
FT CARBOHYD 56 56 POTENTIAL.
SQ SEQUENCE 465 AA; 51798 MW; 29B33FB1EADDCBAD CRC64;

```

Query Match 44.9%; Score 932; DB 1; Length 465;
 Best Local Similarity 39.4%; Pred. No. 2.9e-63;
 Matches 183; Conservative 77; Mismatches 139; Indels 66; Gaps 3;

```

OY 3 VDETLPRKVPGLCARVIALVHRCNTTIAQVNIWTVAVANSTSPQSOLNDSE 62
Db 1 MDNPPSRKPCFCFSRYVLALEMHRCNIVITIAORCSLSTVAVANNITNGSPNTSAE 60
OY 63 -----VLPDPSGSLSKAPKSLP----- 81
Db 61 KRLDMTKNPVYMWSPDVQCIIFSSIFYGAFLIQIVGYSIGYISIKKLIGFLSLSVS 120
OY 81 -----AKSIIIGGFAIWERNGPQPERRLCSIALSGMLGC 117
Db 121 IFIPQAAVGETWIIYICRVQGITQGTVTTAQHEIWMKAPLERGRILSMISLGFILGP 180
OY 118 FPAIIIGFISFTLGMPEVYIFEGVGCYCLLMFVYIDDDPSYMWISTSEKEYITSSL 177
Db 181 FIVLLVGTGICSLGMPWFYIFGIYGVLSLWFLFEDDDKDHVMSSSEKDIYITSSL 240
OY 178 KOOVSSKQPLPIKAMLRSLPIWISICLGFSGHOMLVSTWVVIPIYISSYHVNRDGL 237
Db 241 IQGSSSTRQSLPIKAMLRSLPIWISICLGFSGHOMLVSTWVVIPIYISSYHVNRDGL 300
OY 238 ISALPFIYAWIGVYGLADFLTKKRLIT-VKRIATILGSLSPSALIVSLPYNSGY 296
Db 301 LSLPYLAVIGIAGLQMSDFLRKISITVYKRLFTTSLSPCVITIMCLLYISYSGF 360
OY 297 ITATALLTSLGSLTSCOSGIYINVLDAIPRYSFLMGASRGSSTIAPVYPTVSGFLS 356
Db 361 YSTVIFELIANSTLSFSGQLINLADIAPRYGFLKAVTALIGMGLISLTIAGLILN 420
OY 357 QDPFGMRNVFLLFAVNLGLFLYIFGEADVQEMAKERKILRL 401
Db 421 QDPESWPKIFLMSIINVISIVIFLIRAKAIDQMAKERTL 465

```

RESULT 6

NPT3_HUMAN STANDARD; PRT; 436 AA.

```

AC 000624.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 3 (SODIUM/PHOSPHATE
DE COTRANSPORTER 3) (NA(+)/PI COTRANSPORTER 3).
GN SLIC17A3 OR NPT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA Ruddy D.A., Krommal G.S., Lee V.K., Mintler G.A., Quintana L.,
RA Domingo R., Meyer N.C., Basava A., McGlelland E., Fullan A.,
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harton C., Tsuchihashi Z.,
RA Wolff R.K., Schatzman R.C., Feder J.N.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
CC INTO CELLS VIA NA+ COTRANSPORT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: 091328; AAB82085.1; -
DR EMBL: 090544; AAB53422.1; -
KW Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
KW Sodium transport.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 98 118 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 366 406 POTENTIAL.
FT CARBOHYD 47 47 POTENTIAL.
FT CARBOHYD 56 56 POTENTIAL.
FT CARBOHYD 68 68 POTENTIAL.
FT CARBOHYD 69 69 POTENTIAL.
SQ SEQUENCE 436 AA; 47350 MW; DF02F618E83A572F CRC64;

Query Match 35.8%; Score 744; DB 1; Length 436;
Best Local Similarity 38.5%; Pred. No. 4e-49;

Matches 162; Conservative 56; Mismatches 107; Indels 96; Gaps 5;

OY 10 RAYPSICARVGIYALVHFCNFTTIAQNYIMNTVAMVNSTSPQSLDSEVLPSDF 69
DB 8 RKGPDCSLRYGLALIMHRSNFTMITORVSLAIITAMVNTTQOGLSNASTEGPADAF 67
OY 70 GGLSKPKSLPAKSSIT-----LG----- 88
DB 68 NNSISIKKEPDKASVYQWSPETOGIIFSSINYGILLTLPSCYLAGIFGAKKMLGAGLL 127
OY 88 -----GQAFIMRWGPPOEBSRLCSTATS 111
DB 128 ISSLLFTPLADPFGVILVIMRYQGAQMAATGQFTIMAKMPLERSLITTAGS 167
OY 112 GMLGCFATLLIGFISFELGWPFYIFGVCVCCLLMFVYIDDPVSYPMISTSEKE 171
DB 188 GSAFGSFIILCVGLISQALSMWPFIFYIFGSCVCCLLMFTVYIDDPMHHCISVREKE 247
OY 172 YIISSLKQGVSSKOPLPKAMRLPIWISICGSCSHQWLVSTMVYIPTYISSYTHN 221
DB 248 HILSLAQOPSSPGRAVPKAMVCLPLWALIFGFSHPWLCITLITPTIYSTLLHN 307
OY 232 IRDNGLLSLALPFIYAVNIGWVGYLADFLTKK-FRLITVRIATYI-----L 277
DB 308 IRDSVLSLPLFIAMASCHILGGLADFLSLRLILRIYRKLFSSLDQVSSWESQGL 367
OY 278 GSPSSALLVSLPYLNSGYITATATLLTSCGLTSCQSGIYINVLDIAPRYSSFLMGASR 337
DB 368 GSSQESSLPLDLDSSSVRLISLVGMSFCLQSTCLAMFTSRLD---KQNFKIGPKR 423
OY 338 G 338
DB 424 G 424

RESULT 7
Y006_CAEEL
ID Y006_CAEEL STANDARD: PRT: 576 AA.
AC P34644;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 52.2 KD PROTEIN ZK512.6 IN CHROMOSOME III.
GN ZK512.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Pelodidae; Caenorhabditis.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.

RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Dublin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson P., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.,
RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: 222177; CAA80150.1; -
DR PIR: S40767; S40767.
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
KW Sodium transport.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 382 402 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 438 458 POTENTIAL.
FT TRANSMEM 473 493 POTENTIAL.
SQ SEQUENCE 576 AA; 63108 MW; 71F3A9EFBE5F84CC CRC64;

Query Match 22.5%; Score 466.5; DB 1; Length 576;
Best Local Similarity 32.3%; Pred. No. 4.3e-28;
Matches 101; Conservative 65; Mismatches 134; Indels 13; Gaps 6;

OY 92 IWRMGPPORBSRLCSTALSGMLGCFATLLIGFISFELGWPFYIFGVCVCCLLM 151
DB 198 VWRVAPRPMKRSKLTATATFAGSTAGAVLGPLSAFLVSYSMAPFLVYGVYINATIM 257
OY 152 FVYIDDPVSYPMISTSEKEYYIISLKQVY--SKOP---LPKAMRLPIWISICIG 205
DB 258 FCVTFEKPAPHPPTISOEKIFI---EDAIGHVSNHPTIRSIIPKAIIVSKVYMAIYA 313
OY 206 CFSHOWLVSTMVYIPIYISSYTHVNRDGLSALPFIYAVNIGWVGYLADFLTKK 265
DB 314 NFARSWTFYLLDONOLTYMKLEALGMKIDAGLLAIPHLWGVGVLMGGLADYLSNKKI 373
OY 266 RLIT-VRKIATILGSLPSSALIYSLPYLNSGYITATATLLTSCGLTSCQSGIYINVDI 324
DB 374 LSTTAVRKIPNCGFGEAALFALVATYSD-TTATMALAAAGMSGFAISGFNVNHLDI 432
OY 325 APRYSSFLMGASRGFSSIAPIVPTVSGFLSODPFGWRNVFLLFAVNLGLFLYLF 384
DB 433 APRYATILMGFSNGITGLGTCPTVTEAPTANS-KHGWTSVPLASLIHFTGVTFYAVY 491

OY 385 GADVOEMAKERK 397
 DB 492 ASGLOEMAEKE 504

RESULT 8
 YLID2_CAEEL STANDARD: PRT: 493 AA.

AC 003567;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 52.7 KD PROTEIN C38C10.2 IN CHROMOSOME III.
 GN C38C10.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Johnstone P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
 RA Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulten J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Watson R., Watson A., Wellstock L., Wilkinson-Spoat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP REVISIONS.
 RA Jones S.J.M.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z19153; CAAT9549.1; -
 DR PIR: S28286; S28286.
 DR NORMP: C38C10.2; CE08647.
 KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
 KM Sodium transport.

FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 85 105 POTENTIAL.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 144 164 POTENTIAL.
 FT TRANSMEM 175 195 POTENTIAL.
 FT TRANSMEM 205 225 POTENTIAL.
 FT TRANSMEM 272 292 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 FT TRANSMEM 348 368 POTENTIAL.
 FT TRANSMEM 375 395 POTENTIAL.
 FT TRANSMEM 406 426 POTENTIAL.
 FT TRANSMEM 441 461 POTENTIAL.
 FT TRANSMEM 493 AA; 52737 MW; 60792715D32553DD CRC64;
 SQ SEQUENCE

Query Match 22.1%; Score 459.5; DB 1; Length 493;
 Best Local Similarity 25.7%; Pred. No. 1.2e-27;
 Matches 121; Conservative 82; Mismatches 181; Indels 87; Gaps 11;

OY 6 TLPRKVPSCARSGIALVLFHCNFTIAQNVIMNTVAVANSTSPQSLDSS- 62
 DB 5 TTKPRIVP---STRFALSLVFPCCCLVTYMMRTNMSFAVCNENKTDGVKVSRCG 61
 OY 62 EVLPVDSFGGL-----SKAPSLPAK-----SSILGGQTA----- 92
 DB 62 EMFVEVSNSSVYIGFDMQKQTTGMVLSFFYIGSQIIGHLASRYGRRVFTLLGS 121
 OY 92 -----IMRWGPQERSRLCSIALSGM 113
 DB 122 ALLTLNPNVARTSEVALAILRAIGLQGTFFPAMHTMSWCPLELSLVLCVYTAGA 181
 OY 114 LIGFTALLIGFISE---TLGPFVEYIFEGVCVCCLLMFVYIDDPVSPWISTEK 170
 DB 182 QIGNVYIPLSGELCEYFGDGMPSIFPIIGVFCVLTAVWVYSSDKPATHPATPEEK 241
 OY 171 EYIISLSKQOVSSKQPLP---IKAMRLSPISICLCGSHQWLYSTWVYIPTYISS 226
 DB 242 QYIYAVAVASMGKDTGKVPSTPWIK-ILTSYAVVACWAGHFGAGVATMLVSLPSFLKD 300
 OY 227 YHVHVRDNGLLSALPEIVANVIGWVGYLDELTKK-FRLITVRKATILGSLPSAL 285
 DB 301 YGLMLSSIGAVASIPYATVFLAINAGVLDITLRSKGLISTLITRRAMVALIGGIF 360
 OY 286 IYSLPINSGYTTATATLTLSCG-LSTLCOSGIYINVDIAPRYSSFLMGASRGFSIA 343
 DB 361 LVASGTCGGQ-DVLIYIFITCGMAISGQYAGFVNVLEIAPPSGVTGNTISMA 419
 OY 344 PYIVPVSGLLSQDPENGRVFFLFRVNLGLLFLIEEADVOEMAK 394
 DB 420 GIISPAVSSYLPNPGTOEBMOMVLMITAGILTLGALLESIFASGEVOYPAK 470

RESULT 9
 YRT3_CAEEL STANDARD: PRT: 544 AA.

AC 010046;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 60.5 KD PROTEIN T07A5.3 IN CHROMOSOME III.
 GN T07A5.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Buck D.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z48055; CAAB8134.1; -
 DR NORMP: T07A5.3; CE01648.
 KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
 KM Sodium transport.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 105 125 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 221 241 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.

FT	TRANSMEM	324	344	POTENTIAL.
FT	TRANSMEM	361	381	POTENTIAL.
FT	TRANSMEM	385	405	POTENTIAL.
FT	TRANSMEM	417	437	POTENTIAL.
FT	TRANSMEM	451	471	POTENTIAL.
SO	SEQUENCE	544 AA;	6053 MM;	4615B3C27D9F86CC CRC64;

QY 299 ATATLTSCGLSTLCOSGIYINVL-----DIAPRYSFPLMGASRG-----FSSIAPIV 348
 DB 346 NTTLVMA---LMAIAFFGFGFALGMPVISTAPKE---IVGLCGVFVNFVGNVASTVTP 399
 QY 349 TVSGFLLSODPEFGWRNVEFLFAVNLGLLFYL-IFGEADVDEMAK 394
 DB 400 LVIYGLVSELHSPNALVFFVGCSA--LMAVYCLFVYVDIKRMLQK 444

RESULT 13
 GDDT_BACSU STANDARD: PRT: 455 AA.
 ID AC P42237;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PROBABLE GLUCARATE TRANSPORTER.
 GN YCBE.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE: 95219079.
 RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.:
 RT "Determination of a 21548 bp nucleotide sequence around the 24
 RL degrees region of the Bacillus subtilis chromosome.";
 RL Microbiology 141:269-275(1995).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D30808; BAO06469.1; -;
 DR EMBL: Z99105; CAB12042.1; -;
 DR SUBMITLIST: BG11160; YCBE.
 DR PFAM: PF00083; sugar_tr.1.
 KW Hypothetical protein; Transmembrane; Transport.
 FT TRANSMEM 19 39
 FT TRANSMEM 59 79 POTENTIAL.
 FT TRANSMEM 87 107 POTENTIAL.
 FT TRANSMEM 108 128 POTENTIAL.
 FT TRANSMEM 153 173 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 253 273 POTENTIAL.
 FT TRANSMEM 289 309 POTENTIAL.
 FT TRANSMEM 320 340 POTENTIAL.
 FT TRANSMEM 348 368 POTENTIAL.
 FT TRANSMEM 386 406 POTENTIAL.
 FT TRANSMEM 414 434 POTENTIAL.
 SQ SEQUENCE 455 AA: 49254 MW: A0DFD23F569CC7B8 CRC64:

Query Match 12.2%; Score 252.5; DB 1; Length 455;
 Best Local Similarity 23.8%; Pred. No. 4.4e-12;
 Matches 94; Conservative 74; Mismatches 154; Indels 73; Gaps 16;

QY 19 RVG-----IALVLFHCNFTTIAQN-----VINMTVAWVNSTSPQSQNDSEVL 64
 DB 81 RFGSKITIALSIFWSPFLLTGALIGFSGACTAIIILFLALRFV----- 125
 QY 65 PVDSEGLSKAPKSLPAKSSIIIGGFAIWERMGPPOERSFLCSIALSGMLGCFALLIG 124
 DB 125 -----GLSEAP-SFPGNGRVAS-----WPSSEKGTASAFPSNAGYFAIVIPPLM 170

QY 125 GFISETLGMPVEYIFGVCVCCLLMFVYIDDPVSPYPMISTSEKEYI-----IS---- 176
 DB 171 GMLTFSFGHSHVYVNGIAGILAVIMLKTVY-EPKKHPVNAELAYIQGGGLISMD 229
 QY 176 SLKQVGGSSKQPLPIKAMLSLPIWSICLQCFSHOMLVSTM---VVIPTIYISSYHVN 231
 DB 230 SKSKQFESKMPY-IKOLLTN-----RMILGVIYIAQCITTLTFEPLTFVLYVQARGMS 284
 QY 232 IRDNGLLSALPFIYVAVYVGVGYLADFLTKKFRLLTYKATITLQSLPSSALVSLPY 291
 DB 285 ILEAGFVASLPLALCGFAGVIGVSDILKKRSLSLTFARKVPIIAGMLTSCSMIV-CNY 343
 QY 292 LNSGYITAT--ALLTLTSCGLSTLCOSGIYINVLDIAPRYSFPLMGA-SRGESSIAPVIV 348
 DB 344 TDSAMLVVYVIMSLAFPGKGGALG-----WAVVSDTSPKECAGLSGLFNTFGNIATITTP 399
 QY 349 TVSGFLLSODPEFGWRNVEFLFAVNLGLLFYL 383
 DB 400 IITGIYVNAATGSPFNALVF--VGANMAIATLSYLL 432

RESULT 14
 DGOT_ECOLI STANDARD: PRT: 445 AA.
 ID AC P31457;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE D-GALACTONATE TRANSPORTER.
 GN DGOT.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 93315143.
 RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.:
 RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
 RL genome: organizational symmetry around the origin of replication.";
 RL Genomics 16:551-561(1993).
 RN [2]
 RP SIMILARITY TO DAL5 FAMILY.
 RA Koonin E.V.:
 RL Unpublished observations (OCT-1993).
 CC -1- FUNCTION: INTAKE OF GALACTONATE INTO THE CELL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L10328; AAA62043.1; -;
 DR EMBL: AE000446; AAC76714.1; -;
 DR EMBL: BG11715; DGOT.
 DR PFAM: PF00083; sugar_tr.1.
 KW Transmembrane; Inner membrane; Transport.
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 100 120 POTENTIAL.
 FT TRANSMEM 158 178 POTENTIAL.
 FT TRANSMEM 183 203 POTENTIAL.
 FT TRANSMEM 258 278 POTENTIAL.
 FT TRANSMEM 299 319 POTENTIAL.
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 353 373 POTENTIAL.
 FT TRANSMEM 391 411 POTENTIAL.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:11:04 ; Search time 50.3 Seconds
(without alignments)
552.743 Million cell updates/sec

Title: US-09-391-958-1
Perfect score: 2076
Sequence: 1 MOWDRLIRKVPVSLCSARY.....LIRGADVQZMAKERKLTRL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_protent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1023.5	49.3	497	4	O9Y2C5 homo sapien
2	485.5	23.4	483	5	O61369 drosophila
3	475	22.9	560	11	O62634 rattus norv
4	449	21.6	573	5	O09932 caenorhabdi
5	436	21.0	512	10	O82390 drosophila
6	404.5	19.5	524	5	O94886 arabidopsis
7	388	18.7	413	10	O23065 arabidopsis
8	363.5	17.5	501	5	O16923 caenorhabdi
9	320	15.4	516	5	O10048 caenorhabdi
10	306.5	14.8	499	5	O01791 caenorhabdi
11	304.5	14.7	530	5	O01537 caenorhabdi
12	282.5	13.6	506	5	O23576 caenorhabdi
13	279.5	13.5	420	5	O23558 caenorhabdi
14	279.5	13.5	592	5	O93599 caenorhabdi
15	259	12.5	478	5	O23063 caenorhabdi
16	254.5	12.3	543	5	O44130 caenorhabdi
17	247.5	11.9	516	5	O9XKCI caenorhabdi
18	247	11.9	537	5	O17785 caenorhabdi
19	244	11.8	455	5	O76838 caenorhabdi
20	244	11.8	473	5	O16252 caenorhabdi

21	243	11.7	493	5	O09479 caenorhabdi
22	227.5	11.0	462	5	O76837 caenorhabdi
23	219.5	10.6	419	5	O9XVED caenorhabdi
24	218	10.5	479	5	O23012 caenorhabdi
25	211.5	10.2	485	5	O9XVKO caenorhabdi
26	210	10.1	345	2	P94774 erwinia chr
27	206.5	9.9	493	5	O22832 caenorhabdi
28	205.5	9.9	421	5	O9XUV7 caenorhabdi
29	203.5	9.8	464	5	O09484 caenorhabdi
30	200.5	9.7	466	5	O16976 caenorhabdi
31	198	9.5	552	5	O21538 caenorhabdi
32	192.5	9.3	406	5	O94307 caenorhabdi
33	190.5	9.2	422	2	O34456 bacillus su
34	185	8.9	467	5	O20265 caenorhabdi
35	183	8.8	452	5	O61803 caenorhabdi
36	179	8.6	498	5	O21312 caenorhabdi
37	173	8.3	332	2	O92FRI burkholderi
38	172.5	8.3	452	5	O44595 caenorhabdi
39	172.5	8.3	482	5	O94305 caenorhabdi
40	164.5	7.9	501	11	O9W081 mus musculu
41	164.5	7.9	583	5	O18935 caenorhabdi
42	160.5	7.7	599	5	O22089 caenorhabdi
43	159	7.7	445	5	O45678 caenorhabdi
44	157	7.6	596	5	O18472 caenorhabdi
45	154	7.4	479	5	P90832 caenorhabdi

ALIGNMENTS

RESULT 1
O9Y2C5 PRELIMINARY: PRT: 497 AA.
AC O9Y2C5:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE NA/PO4 TRANSPORTER HOMOLOG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RX MEDLINE: 99253143.
RP SEQUENCE FROM N.A.
RA SHITUI A., TSUNODA T., SEKI N., SUZUKI Y., SUGANE K., SUGANO S.;
RT "Isolation and chromosomal mapping of a novel human gene showing
homology to Na+/PO4 cotransporter.";
RL J. Hum. Genet. 44:190-192(1999).
DR EMBL: AB020527; BAA76653.1; -
SQ SEQUENCE 497 AA; 54055 MW; 128P47B4 CRC32;

Query Match 49.3%; Score 1023.5; DB 4; Length 497;
Best Local Similarity 43.4%; Pred. No. 26-65;
Matches 202; Conservative 70; Mismatches 114; Indels 79; Gaps 2;
OY 16 CSRRYIALVLFHCNFTTIAQNVINITYAMVNSTSPQSLNDSEVLVDFSGLSKA 75
DB 33 CAVRHSLALILDLGNFSIYTOQNLISALPAMVNNTPSPQNASTERPSTDSQYWN 92
OY 76 PSLPLPKS-----
DB 93 LKFRKMAPAYDMSPEITQIILSLNKGSLAPISPGYVAGIRGATYVAGAGFISSFL 152
OY 84 -----SLGGFAIWERNGPQERSRLCSIALSGMLGC 117
DB 153 LFIPLAANGVALLIVIRIVQGIAGVAVLTGQYSIWKNAKPLERSQLTTIAGSGMLGS 212
OY 118 FRIILIGFISELGMPPFYIFIGVGCVCCLLMFVYIYDDPSYTWISSESEYIISL 177
DB 213 FYVLLAGGLLCQRTIGMPYFYIFIGVGCACCPMLPFLITDDPNVHPFISGEXRYIVCS 272
OY 178 KQGVSSKQPLPIKAMRLSPIMWSICLGFSHQWLSTWVYIPTYISSYVHNINDGL 237

Dd	273	AQODCSPGKSLPIRAMIKSLPMLALVSYFCGYWLFYTIMATPTIYSVLQANLRDSGI	332
Oy	238	LSALPFYIAWMYIGMWGYADFLTRK-FRLITVRKIATILGSDPSALLVSLPYLNSGY	296
Dd	333	LSALPFVVGCCIIIGGLADELFLSRKLIRLTIRTKLIFPAIVGLEPPSVYLVLPMWRSSH	392
Oy	297	ITATALLTSCGLTSCSGIYINVDLPAPRSSFFMGMSRGSFSSLAIPVTVSGFILS	356
Dd	393	SMTMFVLVSSAISSFCESGALVNLDPAPRTGTGKLLQVFAHAGASIPFAAGFFIS	452
Oy	357	QDPEFGMRNVFELLFAVNLDGLLFYILEGADVOMAKERKTLRL	401
Dd	453	QDSEFGMRNVFLSAAVNIGSLVFILFGRADVDQWAKKQTFTHL	497
RESULT	2		
ID	061369	PRELIMINARY;	PRT; 483 AA.
AC	061369:		
DT	01-AUG-1998 (TREMBLrel. 07, Created)		
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
DT	01-NOV-1999 (TREMBLrel. 12, last annotation update)		
DE	PURATIVE INORGANIC PHOSPHATE COTRANSPORTER.		
GN	PICOT.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
RN	Ephydroidea; Drosophilidae; Drosophila.		
RA	[1]		
RC	SEQUENCE FROM N.A.		
RL	DA LAGE J-L, ALLAND C ;		
CC	Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.		
DR	-1- SDCCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.		
DR	EMBL; AF024691; AAC39088.1; -		
KW	FLYBASE; FBgn0024472; Dana\Picot.		
FT	Transposrt; Transmembrane.		
FT	TRANSMEM 187 207 POTENTIAL.		
FT	TRANSMEM 292 312 POTENTIAL.		
FT	TRANSMEM 383 403 POTENTIAL.		
FT	TRANSMEM 420 440 POTENTIAL.		
SO	SEQUENCE 483 AA; 52885 MW; EB8E3B7C CRC32;		
Query Match	23.4%; Score 485.5; DB 5; Length 483;		
Best Local Similarity	27.5%, Pred. No. 3,3e-27;		
Matches 123; Conservative	81; Mismatches 163; Indels	81; Gaps	11.
Oy	24	LVIHFNCFTTIQONTYIMNTVMAMNST---SPQSOLD--SSEVLVPVS-----	69
Dd	2	LPLGANAYVMKTN-MSAIYAAMNHRAIKGEERYIDECGDRODIPIDSODGEPPMNA	59
Oy	69	-----EGLSKAPKSLP-AKSSITGGQFA-----IW--	94
Dd	60	ALQGLISSEFYGVYTTPGFILLAKKKGSLEFLGMLINSVPAFLPVARREGGWGL	119
Oy	94	-----ERNGPQENSRCLSIASLGMLCGFTAILIGRFIS--	130
Dd	120	CAVREIOGLGEPVPCPTHAMLAKEWIPENSRMGAAVAAGAFTGISIMPLSGLAERG	179
Oy	130	TLAGPEFYINGGVGCCCLLMFVYIDDPPSYMYISSEKEYIISLKQOVGSSKQTL	188
Dd	180	FDGGMPSIFYGVIGYTWSTAFLLFYEDPSTHKRIDEREKRYINESLIMGDVYIKSPI	239
Oy	189	PIKAMRLSLPIWSICLGCFSHQMLVSTNWVVYPITYISSYHNINRDNGLSALPFIAMV	248
Dd	240	PKFSIVKSILPEFAILFAHHGHNYGETELMTLPTVMKQVLRSLKSNGLSSLPLYAML	299
Oy	249	IGMWGYADFL-LTKKRLLTVRKRIATILGSLPSALLVSLPYLNSGYTTATALLTSC	307
Dd	300	LSMFTSVIADWMISSKRSRLNTRKRIINSIGQYGGLALIASYTGCDRALTIALITIGV	359

Oy	308	GISTLCGGIIIVNDIAPRRSSFLMGASRKFSSIAPIVPTVSGFILS--QDPERG-WR	364
	:	: : : : :	I :
Dd	360	GLNGCIYSGFKRIINHDLIPPRAFGLSTINC SANIAGLAPIAAGNLISPSKPYVMGQ	419
Oy	365	NVEFLLFAVNLLGLEFLYLFECAVOEM	392
	:	: : :	I :
Dd	420	IVFFIAFAYVIICGTFTYNIFESGEROFW	447
RESULT	3		
O62634			
ID	O62634	PRELIMINARY;	PRT; 560 AA.
AC	O62634;		
DT	01-NOV-1996	(TREMBLrel_01, Created)	
DT	01-NOV-1986	(TREMBLrel_01, Last sequence update)	
DT	01-NOV-1998	(TREMBLrel_08, Last annotation update)	
DE	BRAIN SPECIFIC NA+-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
NC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
RN	[1]		
RA	SEQUENCE FROM N.A.S.		
RX	MEDLINE: 94261635.		
RT	NI B., KOSTECK P.R., NADI N.S., PAUL S.M.;		
RT	"Cloning and expression of a cDNA encoding a brain-specific Na(+)-		
RT	dependent inorganic phosphate cotransporter."		
RL	Proc. Natl. Acad. Sci. U.S.A. 91:5607-5611(1994).		
DR	EMBL: U07609; AAA19646.1, -		
SO	SEQUENCE 560 AA; 61665 MW; C7213CFL CRC32;		

Query Match	22.9%	Score 475	DB 11	Length 560
Best Local Similarity	28.0%	Pred. No 2.1e-26		
Matches 111	Conservative 66	Mismatches 156	Indels 62	Gaps 6

QY	12	VSPLCSARYGIALVLFHFCNFTTIAQNVIMNITMAMVNSTSPQSQINDSSEVLVYDSFGG	71
DB <td>161</td> <td>IPSAARVHGCYI-----FVRLDGLVEGYTPAC-----</td> <td>191</td>	161	IPSAARVHGCYI-----FVRLDGLVEGYTPAC-----	191
QY <td>72</td> <td>LSKAPKSLPAKSSILGGGPAIWERMGPOEERSRLCSIALSGMLGCTPALLIGFSIETL</td> <td>131</td>	72	LSKAPKSLPAKSSILGGGPAIWERMGPOEERSRLCSIALSGMLGCTPALLIGFSIETL	131
DB <td>191</td> <td>-----HGIMSKNAFPLEKSRRLTTAFCCGYAQAAYVAMPPLAGLVQIS</td> <td>232</td>	191	-----HGIMSKNAFPLEKSRRLTTAFCCGYAQAAYVAMPPLAGLVQIS	232
QY <td>132</td> <td>GMPEFYVTFGGYGVCCCLMFVYIVDDPVSYSPWISTSEKEYITISLKOQVSS-----</td> <td>185</td>	132	GMPEFYVTFGGYGVCCCLMFVYIVDDPVSYSPWISTSEKEYITISLKOQVSS-----	185
DB <td>233</td> <td>GMSSVFYVYAGSGEIGWYLFWLLVSYESPALHPISIEEEKRY----EDAIGESAKILMPY</td> <td>288</td>	233	GMSSVFYVYAGSGEIGWYLFWLLVSYESPALHPISIEEEKRY----EDAIGESAKILMPY	288
QY <td>185</td> <td>KQPIPIKAMLRSLPIMSICGCFHOMIVSMYVYIPIYISSYVHVNIRDGLSALPF</td> <td>243</td>	185	KQPIPIKAMLRSLPIMSICGCFHOMIVSMYVYIPIYISSYVHVNIRDGLSALPF	243
DB <td>289</td> <td>TKFNTPMKREFTSMKVYIYIVANFCRSMFTFYLLISQPIFEYEEVGFELSKYGLVSALPH</td> <td>348</td>	289	TKFNTPMKREFTSMKVYIYIVANFCRSMFTFYLLISQPIFEYEEVGFELSKYGLVSALPH	348
QY <td>244</td> <td>IVAWYIGWVGYLADFLTLTKRFLTT-VKTIATILIGSLPSSALYISLPLNSGYTATPAL</td> <td>302</td>	244	IVAWYIGWVGYLADFLTLTKRFLTT-VKTIATILIGSLPSSALYISLPLNSGYTATPAL	302
DB <td>349</td> <td>LVMTIIVPIGGIADFLSRHIMSTPTNVRKILNCGGFGMEATLLVYVGSKGV-ATSE</td> <td>407</td>	349	LVMTIIVPIGGIADFLSRHIMSTPTNVRKILNCGGFGMEATLLVYVGSKGV-ATSE	407
QY <td>303</td> <td>LTLSGGLTLCOSGIYINVLDAIPRYSSFLMKAASGCFSSIAVYIPTYSGFLSDPPEG</td> <td>362</td>	303	LTLSGGLTLCOSGIYINVLDAIPRYSSFLMKAASGCFSSIAVYIPTYSGFLSDPPEG	362
DB <td>408</td> <td>LVLAGFGFALISGENVNHLDAIPRYASTILMISNGVGLSCGWCPIYGAMTKHKTRRE</td> <td>467</td>	408	LVLAGFGFALISGENVNHLDAIPRYASTILMISNGVGLSCGWCPIYGAMTKHKTRRE	467
QY <td>363</td> <td>WRNVFFLLFAVNLILGILFYILFGEDADVQEMAKERKIT</td> <td>399</td>	363	WRNVFFLLFAVNLILGILFYILFGEDADVQEMAKERKIT	399
DB <td>468</td> <td>WQYVFLIASIVHYGVIIFYGVAFSEKQPMALPEEMS</td> <td>504</td>	468	WQYVFLIASIVHYGVIIFYGVAFSEKQPMALPEEMS	504

RESULT	4
ID	Q09932
AC	Q09932
DT	01-NOV-1996 (Tremblrel. 01, Created)
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT	01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE	HYDROLYTICAL 63.6 KD PROTEIN K10G9.1 IN CHROMOSOME III.
GN	K10G9.1.
PRELIMINARY:	PRT, 573 AA.

OS Caenorhabditis elegans.
 OC Eukaryota: Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
 OC Rhabditina; Rhabditidae; Rhabditidae; Peleoderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MORTIMORE B., BUCK D.;
 RL Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
 DR EMBL: Z36282; CAAB5289.1; JOINED.
 DR EMBL: Z48055; CAAB5289.1; JOINED.
 DR EMBL: Z48055; CAAB8135.1; JOINED.
 DR EMBL: Z36282; CAAB8135.1; JOINED.
 SQ SEQUENCE 573 AA; 63610 MW; F00BBECC CRC32;

Query Match 21.6%; Score 449; DB 5; Length 573;
 Best Local Similarity 26.7%; Pred. No. 1.5e-24;
 Matches 125; Conservative 84; Mismatches 165; Indels 94; Gaps 14;

OY 1 MOVDELTPRKYPSCSARVIALVLF-----CNFTTAQVNIINIMVAVNST 51
 DB 35 LQEDKKFKLRKV-----RMQIALAHFGPAISFGIRSNF-GVAKRMNIN-----NFT 80
 OY 52 SPQSLND-----SSEVLPLVDS--FGLSK-----APKSLPAKSSILGQFA---- 92
 DB 81 DAYGEVHEKEFPMTGTVEGMMESSFEYGAASQIPAGVIAAKFAPRKLMLGLTFASLLN 140
 OY 92 -----IWERKPPQERSRLCSIALSGML 114
 DB 141 IYTAICLNFHPETDIFVAVVIVQVMOGLALGVCPAMHGVMKRYAPLERSKLATTFPTGAS 200
 OY 115 IGCFAAILIGFISLTLGMPFVFIYFGVGCYCCILMFVYIYDDPSYPMWISSEKEXII 174
 DB 201 VGVWVGLPASALVLFHSFSTPFPVFGALGIWISILMIVSGTSPETHGYISADEKKYI- 260
 OY 175 SSKQOVGS-----SKOPLPIKAMLRSLPIWSICLGFCSHOWLVSIMVVIPTYSIVY 228
 DB 260 ---TEKVGSVANKNMTITLTPWRDMWSTRAVVAIIICSCRSRSPFLILGNOLTYKDWL 316
 OY 229 HVNINDNGLSALPPIVAVVIGVGGYADFLTK-KRPLITVRKIATILGLSPSSALIV 287
 DB 317 HIDIKNSGLIAIFPOLGMCIVTLTSGQLSDYLRSSGKMSTEAVKSVNFEFGVEAVMLG 376
 OY 288 SLPIYNSGVTITALLTSLCGLSTLQSGIYINVDIARYSFSLMGARSGSSINPIVY 347
 DB 377 CLAFPRDPIVATFPLI-TACSGAGAVLGSFNVNHEDIAIRHAPITILMGINGALAGV-G 434
 OY 348 PTVSGFLSQDPEFGMRNVFLLFVAVNLGLFYLIIFGADVOENAKE 395
 DB 435 GIVTNSLTGYQND-GMOWVFLAMSIDIRGITFLIFLPAKGDVLPVARE 481

RESULT 5
 AC 082390 PRELIMINARY; PRT: 512 AA.
 ID 082390;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, last annotation update)
 DE T27A16.25 PROTEIN.
 GN T27A16.25
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsids.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA ROUSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,
 RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;
 RT "Arabidopsis thaliana chromosome II BAC T27A16 genomic sequence."
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL: AC005496; AAC35230.1; -.
 DR PFAM: PF00083; sugar tr. 1.
 SQ SEQUENCE 512 AA; 56498 MW; 933A5143 CRC32;

Query Match 21.0%; Score 436; DB 10; Length 512;
 Best Local Similarity 30.5%; Pred. No. 1.1e-23;
 Matches 101; Conservative 66; Mismatches 140; Indels 52; Gaps 7;

OY 63 VLPLVDSFGLSKAPKSLPAKSSILGQFAIWERKPPQERSRLCSIALSGMLGCTAIL 122
 DB 197 LTVRAFMCVGRG-VAMPAMNNTL-----SKWVYQERSRLALVSGMYGTGTA 248
 OY 123 IGGFISLTLGMPFVFIYFGVGCYCCILMFVYIYDDPSYPMWISSEKEXIISLQOYG 182
 DB 249 FSEFLIHQFGWSPVFSFSLGTWMLTLMTAESSPLDEPDLPEERKLIADNC----- 304
 OY 183 SSKOP---LPKAMLRSLPIWSICLGFCSHOWLVSIMVVIPTYSIVVHNINDNGLS 239
 DB 304 ASKEPKYSIPKRLILSKPPVVALISCHFNWCTFTLTMPTIYHOVLKFNLMESGLLS 363
 OY 240 ALPEIYAVVIGVGYLADFLTKRPLITVRKIATILGLSPSSALIVSLPIYNSGYTA 299
 DB 364 VEPWMTMAISANAGMIADTLVSRGFSVTVNRKIMQITIGLGPAPFLTOLKHIDS---PT 420
 OY 300 TALLTSLC---GLSTLQSGIYINVDIARYSFSLMGARSGSSINPIVYPTVSGFLSQ 357
 DB 421 MAVLCACSGQGDPAFSQSGIYSNHDIPAPYSCVLLGLSNFAGVLAGVLTATGHIH 480
 OY 358 DPEFGMRNVFLLFVAVNLGLFYLIIFGAD 388
 DB 481 G---SMDDVFTISVGLYGVITMNLFSIGE 508

RESULT 6
 AC 094886 PRELIMINARY; PRT: 524 AA.
 ID 094886;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, last sequence update)
 DT 01-AUG-1998 (TREMBlrel. 07, last annotation update)
 DE NA(+)-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
 GN NAPI-T.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-OVERLAIN;
 RA MACIVER B.;
 RA Thesis (1997), University of Edinburgh, Edinburgh, UK.
 DR EMBL: Y07720; CAAB6892.1; -.
 DR FLYBASE: FBgn0016684; NAPI-T.
 SQ SEQUENCE 524 AA; 58814 MW; 3F888436 CRC32;

Query Match 19.5%; Score 404.5; DB 5; Length 524;
 Best Local Similarity 31.6%; Pred. No. 2e-21;
 Matches 95; Conservative 60; Mismatches 141; Indels 5; Gaps 5;

OY 95 RMPQPPERSRLCSIALSGMLGCTAILIGFISLTLGMPFVFIYFGVGCYCCILMFVY 154
 DB 170 KWIIPERRSFVS-AVLGSSGVALLFPYFIIDTTRMWMVYVIGIYGTLMFIWQFL 228
 OY 155 IYDDPSYPMWISSEKEXIISLQOYSGSKOPLPIKAMLRSLPIWSICLGFCSHOWLVS 214
 DB 229 VEPSPAEHRILADSEKFKFLKSIGASTGSKGTPPKAITSPPWLNVVAQMGKTLGFL 288
 OY 215 TAVVYIPTYIISVYVHNINDNGLSALPPIVAVVIGVGYLADFL-TKFRPLITVRKI 273
 DB 289 TLMTHTPTYFRLIHHNINIRATGSLGPLHMRLLFAVVSIFADVILRTDKMSRTVNRKL 348

OY	274	ATTTGSEFPSSALIVSLPYLNTSGYTITATTALLTTLSCGLSTLCOSGITIINVDIAPRRSSFLM	333
Dd	349	ATFTCCGTGGKGLVLAFAFGYNATAIVLVYATMLHGAVSSGPILASWDLSPNAGIVL	408
OY	334	GASRQFSSIIAPIYPTVTSVGFFLLSDPDEF-GWRNVPELLFAVVLLGL-LFYLIIFGEADVOE	391
Dd	409	GVSNGIGMGPGETISFEIFYGQLTHNNQTIDAKNV-FLLTSLMTLSGILTYLFESKLDP	467
OY	392	W 392	
Dd	468	W 468	
RESULT	7		
O23065			
ID	O23065	PRELIMINARY;	PRT; 413 AA.
AC	O23065;		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-AUG-1998 (TREMBLrel. 07, Last annotation update)		
DE	BAC IG005110.		
GN	A_IG005110.NN.		
OC	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;		
OC	core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;		
OC	Arabidopsits.		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN-CV. COLUMBIA;		
RA	ANDREWS S.;		
RL	Submitted (JUL-1997) to the EMBL/genBank/DBJ databases.		
RN	[2]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN-CV. COLUMBIA;		
RA	WATERSTON R.;		
RL	Submitted (JUL-1997) to the EMBL/genBank/DBJ databases.		
EMBL:	AF013293; AAB62846.1; -.		
OR	SEQUENCE 413 AA; 45791 MW; 0D120117 CRC32;		

	Query Match	Similarity	18.7%	Score 388;	DB 10;	Length 413;
	Best Local	Similarity	25.7%	Pred. No. 2,4e-20;		
	Matches 106;	Conservative	69;	Mismatches 133;	Indels 104;	Gaps 11.
QY	19	RGIALVLHFCFETTTIAQNVINNTIMVAVNVTSPQSQ-LNDSSEVLPPY-----	67			
DB	60	RMWIVYLCCFSF-----LTCNMORVNMSIALILPMSQCYNMSSATVGLIQSFFEWGYLL	113			
QY	67	-----DSFGGLSKAPKSLPAKSSILIGGFALWERKGPPOEHSRLCSIALSGMLCC	117			
DB	114	TQILIGIWADKRGKGVA--MPANNMML-----SKMIPVSESRSLATLYSGCYTLGS	163			
QY	118	FTAILIGFISFTLMEPFVYIFGCGVCCLLMFVYIDDPVSPWMISTEKEYITSL	177			
DB	164	VTGLAFSPMLITKRPMPSPVSYFSGSLGSIWFLMLKFAIASSKDDPDLSEKKYVLGS	223			
QY	178	KQOVSSKOP--LPKAMLRSLPIWISICLGFSSHOWLVSTVWYIPTY-----T	224			
DB	224	K-----PREPVVIVIPWKLILSKRPWALIIHSFCHWNGFTILLTWPPTYNOARSSASYI	278			
QY	225	SSVHV-----NIRONGILSLALPEFIYAWYIGWVGVGLADEFLTKRRLITVKRIATI	276			
DB	279	SLTFNIFCEQVILKFNLTESGLLCVLPWLTMAAFANIGMIADTLVSRG-----	327			
QY	277	LGSLPSSALIVSLPYLNSGYIFATALLTLSCGLSTLCQSGIYINVDIAPRSSFLMGAS	336			
DB	327	-----LSTITNSDAFSSQSGIKLSSNHQDIGPRAGVYLLGS	360			
QY	337	RGFSSIAPIYIVTVSGFLLSQDPBFGKRWNVFLFLPAVNILLGLFTYLISEAD	388			
DB	361	NTAGVLAGVGTAAATGYILQRG---SMDVYFVAVVALYILGLVNMVLTFTGE	409			

RESULT	8			
ID	016923	PRELIMINARY;	PRT;	501 AA.
AC	016923;			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)			
DE	F25G6.7 PROTEIN.			
GN	F25G6.7.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;			
CC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.			
LN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., MEDLINE; 94150718.			
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,			
RA	BOURFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,			
RA	CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,			
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,			
RA	JONES M., KESHAM J., KIRSTEN T., LAISTER N., LATREILLE P.,			
RA	LIGHTNING J., LOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,			
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SANDERS D., SHOWNKEEN R.,			
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,			
RA	THERRER-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,			
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans."			
RT	Nature 368:32-38(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	NELSON J., WOHLDMANN P.,			
RL	Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	WATERSTON R.,			
RL	Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AF022973; AAC25800.1; -.			
DR	SEQUENCE 501 AA; 55339 MB; 2D9C2975 CRC32;			

Query Match	17.5%	Score 363.5;	DB 5;	Length 501;
Best Local Similarity	23.7%;	Pred. No. 1.6e-18;		
Matches 108;	Conservative 98;	Mismatches 165;	Indels 85;	Gaps 13;
QY	17	SARCIATLVHFCNFTTTAQNINMITHVAMVNS-----ISPSQ---	LNDSE-----	63
		: : : :	: : : :	: : : :
Db	25	SWRFMTSITMLCFCCFCVHLMNSNMGMATVCMVSATYDNETYYPENAPPLDMSSDEQGY		84
QY	63	-----VLPSDRSGIS-----	-----KAPKS---	80
		: : : :	: : : :	: : : :
Db	85	IFSAPNAGLIVLFTGGMADKFNATYMTLLVSAIASLANFLTPMNAPITSYVAIFSRFLV		144
QY	80	-----PAKSIILGGGFALIMERGMGPQERSRLCSIALSGMLLCCFATAIL---	GGFIS 128	
		: : : :	: : : :	: : : :
Db	145	GFADALLQPPAMSLI-----TRNFPITSERSYALGLTGGQIOI--TIIITIPAAALCS		195
QY	129	ETL---GMPFVYIFGGVGCVCCLMFVYITDDPVSYPMISTSEKEYI--ISSLKQVGS		184
		: : : :	: : : :	: : : :
Db	196	QETIRGGMSITFYLSGFIQVLFISFYIFLGDADPKSCQICISDNELKFTIISNQSDEVGAK		255
QY	184	-SKOPLPIKAMLRSPISISICLGFCSHOWIVSTWVYIPTVYSYVHNIRNGLLSALP		242
		: : : :	: : : :	: : : :
Db	256	KTERVYPMKNIILKSAVNAVSIVLVCHEPPLMTLLMFPLSYLHDVHHHSTNGLLSALP		315
QY	243	PIVAVVIGVAVGVLADFL--LTKKFRLITVRIKATILIGSLPSSALLVSLPLNSGYI--TAT		300
		: : : :	: : : :	: : : :
Db	316	TVSLMFAKIGSSLYMTWILQKNTTWKDKTICKVLNIGSIGLGVFLLAATFLDNEHAMAV		375
QY	301	ALLTSCGISTLCQSGIYINVLDIAPRYSSFLMGASRGFSSIAPIVPTVSGFLLSODPE		360

Db 376 LFLCISMASGLHPPGCOLALVSAPRASCAGVCTFEFFAASGIIHPITTKMYKDRTA 433

QY 361 FGRWVEFFLEFAVNLGLLEFYLLIGEADVOEWAKER 396

Db 436 AENMLVFIYSTIVIAIFPIIIFINWGSIEVOMAKSK 471

RESULT	9	
Q10048		
ID	Q10048	PRELIMINARY;
		PRT;
		516 AA

DT 01-JAN-1999 (TEMBLrel. 09, Created)
DT 01-JAN-1999 (TEMBLrel. 09, Last sequence update)
DT 01-JAN-1999 (TEMBLrel. 09, Last annotation update)
DE HYPOTHEetical 56.0 KD PROTEIN T09B9.2 IN CHROMOSOME X
DE map09.2

GN 10789.2.
OS *Caenorhabditis elegans*.

0C Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditida;
0C Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis

RN [1]
RP SEQUENCE FROM N.A.
R1 REDUCED T

Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases:
 RA KERSHAW J.;
 RL
 CC

CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL)
CC -1 SIMILARITY: TO SODIUM/PHOSPHATE COTRANSPORTER.
DR EMBL: Z47470. CAA87340.1. -

DR PEAM; PF00083; sugar_tf: 1.
DR WORMPEP; T09B9.2; CE01650.
DR EMBL; E07000; CEM07510:1;

KW	Hypothetical protein; Transmembrane.
FT	TRANSMEM 22 42 POTENTIAL

FT	TRANSMEM	120	140	POTENTIAL
FT	TRANSMEM	186	206	POTENTIAL

FT	TRANSMEM	208	228	POTENTIAL
FT	TRANSMEM	276	296	POTENTIAL
FT	TRANSMEM	304	324	POTENTIAL

FT	TRANSMEM	377	397	POTENTIAL
FT	TRANSMEM	351	371	POTENTIAL
F1	IKANSMEM	304	324	POTENTIAL

FT	TRANSMEM	400	420	POTENTIAL
FT	TRANSMEM	445	465	POTENTIAL

SQ SEQUENCE 516 AA; 55990 MW; F6DB2754 CRC32;

Query Match	15.4%	Score 320	DB 5	Length 516
Best Local Similarity	25.4%	Pred. NO. 2e-15		
Matches 105	Conservative 78	Mismatches 177	Indels 54	Gaps 13

Dd
oy
29 CN-----FTTAAQNVLA--MNTITVMAVMVNSTSPQSOLD-----SSEILEPV 66
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
71 CNIIPKGFAMSSIQGLGIYSGNGFSLFMVITTGQADRLNGKWTIVAAAFIYSNAVLPI 130

Db 131 SAGASEALVEFLRVLTGFGDALIS-PASSSLI-----TRWPPKRRPSALGIIVTSQRQ 182

```

Db 183 IGTLLIPICGWLCCSDGSKFLGWPALFYLLSSVVAALVIVWVFSADKPSKHLCSHN 242

```

```

0v      226  SVYHVNIPBNCILSADETVAUWICMGCVLADLITKYE-BITWVDTATTICSTDCSA 284
Db      243  EEAYINRKIEEENIGKRRNNRKPTPKAIFTSKQVVAVALVCHEPPLVIMLQFLPKFFS 302

```

Db	303	DVLCISNTVNGVLVSALPMALICLSKCLASLASLYLTANGVLRKTKSCIKFNFIASLGIGI	366
Qy	285	LIVSLPYLNS--GYITATALLTSCGISTLCQSGIYINVDIAPRRSSFLMKASRGFFSI	342

Db 363 CIAATPLMSNQHAITNAIIILCLANAFAGLHTPCVLTATVQLAPAESGITITGLAFVASC 422

Qy 343 APVIVPTVSGFLLSQDPEFGWRVVEFLLEAVNLGLFLFYITGEADVQEWAKER 396

Db 423 FSIENKLLISQILRTGSKHEWTIVFEISAFVALPTIFFTLNGSAERTEMASNR 476

RESULT 10

ID	001791	PRELIMINARY;	PRT;	499	AA.
AC	001791;				

DT	01-JUL-1997	(TREMBlrel. 04, Last sequence update)
DT	01-NOV-1998	(TREMBlrel. 08, Last annotation update)

DE
GN
SIMILAR TO SODIUM DEPENDENT-PROSPHATE TRANSPORTER.
F12B6.2.

OC Eukaryota; Metazoa; Secernentea; Rhabdilitia; Rhabdilitida;
OC Rhabdilitina: Rhabdilitidae: Peloderinae: Caenorhabdittis

SEQUENCE FROM N. A.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RX MEDLINE; 94150718.

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
BA CARDNER A. GREEN D. HARRING T. HITT I. TIED M. TUNCKSON T.

RA JONES M., KERSHAW J., KIRSTENJ., LAISTER N., LAURELLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA

RA SMALDON N., SMITH A., SONNHAMMER E., SPADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAIDIN M., VAUGHAN K., WATERSTON B.

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C

RL	Nature 368:32-38(1994).
RN	[2]

RC STRAIN=BRISTOL N2;
PA DATEEV 1 M00CT 1

RN [3] submitted (Mar. 1997) to the EMBL/GenBank/DBJ databases.

RC STRAIN-BRISTOL N2;
RA WATERSTON R.;

DR EMBL; AF003138; AAB54154.1; -.

SEQUENCE 499 AA; 54949 MW; 39AED/31 CRC32;

Query Match	14.8%	Score 306.5	DB 5	Length 499
Best Local Similarity	24.5%	Pred. No. 1.8e-14		
Matches	80	Conservative	72	Mismatches 134
				Indels 41
				Gaps 7

```

0Y  LPAKSLIGGFAIWERMGPPQERSKLSIALSMLLGFYALLIGGISETL----- 132
    :| | :| | | | :| | :| | :| | :| |
166 IPAGSVLISS-----WFLPEKSTAMAIFTTGNOIG-----IANSMFLTAKLCQLHFE 214
Db

```

Db 215 GMPLEFIVYGLIGAVFLIMHVRLADKPRESKYITATELYI-----KGGKQRNRRAE 267

Db 268 TIVRATPYMKITLNGCYCAICACSFASQSVLVALVYLPKRYNQIAFKMNLTHNGINSSLP 327

```

Db      328  FFQMITKLLFAIADKVKQRKVATAVTKVSNIAISFASAFIVIAAY---GPFDSAEI 384
Ov      303  LTIAGGISTICQS---GIYINVIDIAPRYSFLMGASRGSSIAIVTPVTYSGFILLSD 358

```

Db 385 VOLSIIVSMAFSAVYPGYNSTIVTAPQFTAFISSYAQYQIASTLAPIVIGRITSHG 444

QY 359 PEFGRNRFVFLFAVNLGLFLYIFG 385

Db 445 TTYEKCAPYSLAGVLAVTGLIFQIFGG 471

RESULT 11

ID 001537 PRELIMINARY; PRT: 530 AA.

AC 001537;

DT 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE COSMID F21F8.

DE F21F8.11.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COUTSON A., CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GRADNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATEILLE P., LIGHTING J., LOYD C., MCMURRAY A., MONTMORE B., O'CALLAGHAN M., PARSONS J., PERCY C., RIKKEN L., ROOPE A., SANDERS D., SHONKKEEN R., SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STILSTON J., THIERRY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA WILSON R., FAVELLO A., LE T.T.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: D97000; AAC47997.1; -.

DR PFRAM; PF00083; sugar_trt.1.

SQ SEQUENCE 530 AA; 58783 MW; 6C36C5A0 CRC32;

Query Match 14.7%; Score 304.5; DB 5; Length 530;

Best Local Similarity 22.9%; Pred. NO.2.6e-14;

Matches 110; Conservative 69; Mismatches 186; Indels 115; Gaps 15.

QY 15 LCSARYGIALVYL-FCNFTTIGNQINYM-----TNYAVNVSPOSOLNS 60

DB 38 MCOMAHIGHLSLSCNCNSTAVA--LMTNNAITVEGTESTILMSILEKTSSEQLGG 94

QY 61 S-----EVLVDVSFGG----- 72

DB 95 DDGESCRTKEGVKIDYGGTFIMSYWQGIYSAFLGGFIPSPAGVLVDRFSARHILS 154

QY 72 -----LSKAPKSLPAKSSILG--GGFA-----IWEWGPQERSRLC 106

DB 155 VALMLTTLASLMLPVLSTIYIGEGKAFAGFYVGISETMLIPISNMYTKWIPDKINSLSAA 214

QY 107 STALSG-MLLGGFTALILGGFSETLGMPEFVFYIFGVCVCCLLMFVVIYDDPSYPMI 165

DB 215 SVFTAGNQLSGFNGHILYAEIACASSFGSSITYSASLFGISMLVLMHLTVRNSPHNTKWI 274

QY 166 STSEKEYITISLKQGVSS-KQPLDIPKALMLRSPLWISICLGFCSHQWLVTMYVIPTVI 224

DB 275 HKRELDYLIANNIPRHPBSVYKTKTPWRDLTKSVFMSLFNSVNGMMALILFYVPIYE 334

QY 225 SSYYHVNIRDNKLLSALDPPIYAVWIGWGGYIADPLLTITKF-----RLITVKIATILGSL 280

DB 335 KDVLMLDVQSNNGFYSAIPISMLAKILMGYILMDKRNKKIISPATYKLSOFASMG-- 393

QY 281 PSSALIVSLPYINSGYITAT--ALLTILSC-----GLSTLQSGYIINVLDAIPRYSPL 332

DB 393 -ISVSCFLRTYNN---CATPPYALVLLSSVSAFGLST--SGFTSLISTAPSHIGTL 444

RESULT	ID	PRELIMINARY:	PRF:	506 AA.				
023576	023576							
AC	023576							
DT	01-NOV-1996	(TREMBLrel. 01, Created)						
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)						
DT	01-NOV-1998	(TREMBLrel. 08, Last annotation update)						
DE	SIMILAR TO SODIUM/PHOSPHATE TRANSPORTER. NCBI GI: 1086815.							
GN	ZK682.2							
OS	Caenorhabditis elegans.							
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;							
OC	Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Pelodierinae; Caenorhabdilitis.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RX	MEDLINE: 94150718.							
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,							
RA	BOFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., FULTON L.,							
RA	CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,							
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,							
RA	JONES M., KESHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,							
RA	LIGHTING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,							
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOMNEEN R.,							
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,							
RA	THERRIET-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,							
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROUT J., WOHLIDMAN P.,							
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.							
RT	elegans."							
RL	Nature 368:32-38(1994).							
RN	[2]							
RA	SEQUENCE FROM N.A.							
RA	DU Z., LE T.,							
RL	Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.							
RP	[3]							
RA	SEQUENCE FROM N.A.							
RA	WATERSTON R.,							
RA	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.							
DR	EMBL: U4110; AAA82415.1; -.							
DR	PFAM: PF00083; sugar_tf; 1.							
SO	SEQUENCE 506 AA; 55385 MM; 8B04E7F9 CRC32;							
Query Match 13.6%, Score 282.5; DB 5; Length 506;								
Best Local Similarity 24.6%, Pred. No. 9.2e-13;								
Matches 82; Conservative 70; Mismatches 140; Indels 41; Gaps 11.								
QY	85 ILG-GQFAIW-----BRMGPOEQRSLRSLGSLGCGTALIGGFISETL----- 132							
DB	158 VMGFGQGVLMPLMOMLVLMQWPEVNEKSTALATATGNNL----SVIAMFAETAEICQLPW 213							
QY	132 GMPFVFYIFGGVGCYCDLMPFVVIYDDPV-SYPMWISTEKEYIISLKQOVSSKQPLPI 190							
DB	214 GWPMAFHYAVVCGIYMCITWYMYVDSCHADKRLSDELAHYITP---ERVRLRQHPRW 270							
QY	191 KAMLRSLPIWISLCGCFESHQMLVITWVVIYIPYIISYVHNVRDNGLLSALPFIYAWYIG 250							
DB	271 MALMKSPVVMVMAIASSFAHNVYTVGTITLYLPIYTVVLMSTJSGMLNSALPFIQLISK 330							
QY	251 MVGGIADPFLTLTKRRLITVRKIATILGSLPSALIVSLPYL-----NSGYTATALLT 304							
DB	331 VFYAGMAE--SARRKDWMDINRITFCNSSASFGAIGALICGLCLDCCSGR-----AATF 383							
QY	305 LSCGLSTLCQSGIYI-----NVLDIAPRYSSFLMGASRGFSSIAIVYVFTVSGFLLSDP 359							
DB	384 LIC--LAMCFVSGYIPGNTSAVTITAPQGTAAIAAFSRWQIASSVAPYHIGAVTKQCT 441							
QY	360 EFGMRNVFPLFAVNLGLIFLFIIGEADVDQM 392							

Db 442 ADEWKIVEAVIAIAICVYTGIFFOCGGTASLQDW 474

RESULT 13

ID Q23558 PRELIMINARY; PRT; 420 AA.

AC Q23558;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE ELEGANS COSMID ZK652.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

CC Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2; TISSUE-WHOLE ANIMAL;

RX MEDLINE; 92168156.

RA SULSTON J., DU Z., THOMAS K., WILSON R., HILLIER L., STADEN R.,

RA HALLOAN N., GREEN P., THIERRY-MIEG J., QIU L., DEAR S., COULSON A.,

RA CRAXTON M., DUREIN R., BERKS M., METZSTEIN M., HAWKINS T.,

RA AINSCOUGH R., WATERSTON R.;

RT "The C. elegans genome sequencing project: a beginning.";

RL Nature 356:37-41(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2; TISSUE-WHOLE ANIMAL;

RA DU Z.;

RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.

DR EMBL; L14429; AAA28211.1; -

SEQUENCE 420 AA; 46281 MW; 4E12C821 CRC32;

Query Match 13.5%; Score 279.5; DB 5; Length 420;

Best Local Similarity 24.5%; Pred. No. 1.2e-12;

Matches 79; Conservative 62; Mismatches 166; Indels 15; Gaps 7;

QY 80 PAKSIIIGGQFAIWERGPPQERSRLCSI-ALSGMLGCFATALLIGRFISLTGMPVFEY 138

Db 5 PTNNATITG-----NMFPSSSEKSTALSTFTLGNQIASAGSPMAAANCASDLGMPATFY 57

QY 139 IFGGVGVCCLLMFVVIYDDPVSPWISTSEKEYIISLKOOGVSSK--QPLPIKAMTRS 196

Db 58 FAGIFATGWSLIMFPTTASHPAKYMMTKKEKEYLLANVVKVHSEKTRSLIPYSKILTS 117

QY 197 LPISICLGC-FSHOWLSTWVVIPIPIYISSVYHVNINDGLSLAPITVAMVIGVGY 255

Db 118 -PAFLGQLQCHFEFVLFMTLQIYLPSEYKEVHLGLVANGTFEAIPIPMIFNVWGI 176

QY 256 LADFLTKKFKRLIT-VKRIATILGSLPSSALIVSLPIYNSGITRTALLTISCGISTL-- 313

Db 177 GIDKIKENKILSNKAVVSHGVASFSSFLILALFVDCSNPTTGLIFCLAMTSMGT 236

QY 313 CQSGIYINVLDIAPRYSFSLGASRGFSIAPIVPTVSGFLLSODPFGMRNVFELFA 372

Db 237 FVSGGYTSLSLAPQYITMNAISMFVAMIGRLITTPAVMSFRKDKGTAAEMONITGCSL 296

QY 373 VNLGLFLFYLIIFGEADVQEMAK 394

Db 297 AHIFSGSIFLFGSGELQDMAK 318

RESULT 14

ID Q93599 PRELIMINARY; PRT; 592 AA.

AC Q93599;

DT 01-FEB-1997 (TREMBLrel. 02, Created)

DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)

DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)

DE T28F3.4 PROTEIN.

GN T28F3.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

CC Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RA MORTIMORE B.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z82285; CAB05296.1; -

PFAM; PF00083; sugar-tri.1.

SEQUENCE 592 AA; 65489 MW; 46C99E3B CRC32;

Query Match 13.5%; Score 279.5; DB 5; Length 592;

Best Local Similarity 22.5%; Pred. No. 1.8e-12;

Matches 81; Conservative 74; Mismatches 146; Indels 59; Gaps 10;

QY 79 LPAKSIIGGQFAIWERGPPQERSRLCSIALSG-MLGCFATALLIGRFISLTGMPVFE 137

Db 190 LPANNAITIA-----NMFPSEKSTALSTFTTGNQAGAGNPVAAASICASSFGMPSTF 242

QY 138 YIFGGVGVCCLLMFVVIYDDPVSPWISTSEKEYIISLKOOGVSSKQPL-----PIK 191

Db 243 YFASITVSTWVCMFLTASNOPSKCKVMKTERDYLDANVARSNKTNVSLRSLVPPYS 302

QY 192 AMLRSLPTWISICLGCFSHOWLSTWVVIPIPIYISSVYHVNINDGLSLAPITVAMVIGM 251

Db 303 KILKSPAPFLAQLLCOFVINFITLLQIYLPNFKVHLHGLVANGTYSVPIVPAFKI 362

QY 252 VGGYIADFLINK-----RLITVRKIATILGSLPSSALIVSLPYL-----NSGYTAT 300

Db 363 VVGITIDAKKEKKIISGTFGVLSOSIANFGAL-----FLILITFYVDCNTPTLGFV--- 416

QY 301 ALLTISCGISTL-----CQ-----SGIYINVLDIAPRYSFSLGASRGFSI 342

Db 416 -----FFCMAISAIHSAHPINTENSRTGCMGTLVSGYTSLSLAPITVAMVIGVCAAM 471

QY 343 APVIPTVSGFLLSODPFGMRNVFELFAVNLGLFLFYLIIFGEADVQEM-----AKERK 397

Db 472 GRLATPATVGLIKKQNTISEMOTLFLVCAANILIGAVFLVGSGLQEMGLDEDAKEMK 531

RESULT 15

ID Q23063 PRELIMINARY; PRT; 478 AA.

AC Q23063;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE SIMILAR TO SODIUM-DEPENDENT PHOSPHATE COTRANSPORTERS.

GN T28H11.8.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

CC Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DUREIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SANDERS D., SHOWNKEEN R.,

RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SOLSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMANN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";

RL Nature 368:32-36(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA NELSON J., WOHLDMANN P.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

Align seg 1/1 to reverse of: AW261723 from: 1 to: 685

```

208 SerHisGlnTrpLeuValSerThrMetValValTyrIleProThyTyrI 224
    ::::::::::::::::::::| | | | | | | | | | | | | | | | | |
685 ACCACACAGCTGCGCATGTTAACACCTTTATTAATGACCTCCCACTACAT 636
224 eSerSerValTyrHisValAsnIleArgSpAsnGlyLeuSerAlaL 241
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
635 CAGTTCTGATTCAGAAAGTTACATCAGACACATGAGTGGTCCGTCTCTC 586
241 eubProheleValAlaTrpValIleGlyMetValGlyGlyTyrLeuAla 257
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
585 TTCCCTTATTTGTTGCGTGGGTCCTTGATCTCCAGCGCGGCGGCA 536
258 AspPheLeuLeuThrLysLysPheArgLeuIleThrValArgLysIleAl 274
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
535 GATTTCTACTGAGACAGAAATTTAGGCTCATACCTGAGAAAATTCAT 486
274 aThrIleLeuGlySerLeuProSerSerAlaLeuIleValSerLeuProT 291
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
485 CACACTTTTGAAGAAATGCTCTCCAGCGCTCGTTGGGCTCTCCCT 436
291 yrlLeuAsnSerGlyTyrIleThrAlaThrAlaLeuThrLeuSerCys 307
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
435 ATATCCATCCAGCTATATTAACAACAATTTATTTCTGACATTTCTCTG 386
308 GlyLeuSerThrLeuCysGlnSerGlyIleTyrIleAsnValLeuAspI 324
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
385 GCACTGTCCTCTATCTCAGCGCAGAAATCATATTAATGCGTTAGACAT 336
324 eAlaProArgTyrSerSerPheLeuMetGlyAlaSerArgLysPheSer 341
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
335 TGCTCCAGGTAATCCAGCTTCTCATGCGAACATCAAGAGGATGGCAC 286
341 eTrIleAlaProValIleValProThrValSerGlyPheLeuSerGln 357
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
285 ATTCATCGCGTGTGCTGACCATGTTGCTGCGCTTTTCCCTCAGCAG 236
358 AspProGluPheGlyTyrPArgAsnValPhePheLeuLeuPheAlaVal 374
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
235 GACCTGTAATCGGGTGGAGAAATTTCTTTTGTAGTGTTCAGATTAA 186
374 nLeuLeuGlyLeuLeuPheTyrIleIlePheGlyGluAlaAspValGln 391
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
185 CCCACTGTGCTTAATCATCTACCTGCTGTGGGAAAGCAGATGTTCAAG 136
391 IuTrpAlaLysGluArgLysLeuThrArgLeu 401
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
135 AATGGGCTAGAGAGAGAAAGCTCACTCGTTG 104
seq_name: gb_est29:AI552795
seq_documentation_block:
LOCUS AI552795 641 bp mRNA EST 23-MAR-1999
DEFINITION mm25d12.y1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:522551.5' similar to SW:NP04_HUMAN 000476 SODIUM-DEPENDENT
PHOSPHATE TRANSPORT PROTEIN 4 ; , mRNA sequence.
ACCESSION AI552795
VERSION AI552795.1 GI:4485158
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 641)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr,K.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterson,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)

```

COMMENT

On Mar 16, 1998 this sequence version replaced g1:2961894.
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40RP from Gibco
 High quality sequence stop: 397.

FEATURES

source

```

1..641
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:522551"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/cisue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (Kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GATTCGGCAGAG 3' -3' adaptor
sequence: 5' CTCGACTTTTCTTTTCTTTTCTTTT 3'"
BASE COUNT 204 a 138 c 146 g 151 t 2 others
ORIGIN

```

alignment_scores:

```

Quality: 602.50 Length: 182
Ratio: 3.766 Gaps: 2
Percent Similarity: 87.912 Percent Identity: 64.286

```

alignment_block:

US-09-391-958-1 x AI552795/rev ..

Align seg 1/1 to reverse of: AI552795 from: 1 to: 641

```

221 ProThrTyrIleSerSerValTyrHisValAsnIleArgSpAsnGlyLe 237
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
641 CCAACTACATCAGTCTGTGATCAAGATTACAT...CAAGACACTATGGGT 595
237 uLeuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 253
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
594 TCTGTCTTCTCTCTCTCTTATTTAGTGGCTGGCTTGTGTCTGTGGA 545
254 GlyTyrLeuAlaAspPheLeuLeuThrLysLysPheArgLeuIleThrVa 270
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
544 GCGTGGTGACATTTTCTACAGACAGAAATTTAGCGCTATTAACGT 495
270 IArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeuIleY 287
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
494 GAGAAATTCATCATCACTTTTAGGAAATGCTCTCGACGCCCTGTGTG 445
287 aLseLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAlaLeu 303
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
444 CGGCTGCGCTTATTCATCAATCCAGCTATTTACAAACATTTATTTCTG 395
304 ThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGlyTyrIleAs 320
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
394 ACAATTTCTGTGACAGTGTGCGCTCATCTCAGCGCAGGAAATCATTA 345
320 nValLeuAspIleAlaProArgTyrSerSerPheLeuMetGlyAlaSer 337
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
344 TGCCTTAGACATTTGCTCCAAAGTATGCGACCTTTCTCATGGAACATCA 295

```

```
337 rglghpseraserllealaprovalillevalprothrvalserglyphe 353
|||||.....:|||||.....:|||||.....:|||||.....:
294 GAGGATTGGACATTCATCGCTGCTGCTACCATTTGCTGCTTT 245
354 LeuleSerGlnAspProglupheglytrparGAsNValPhepheleule 370
|||||.....:|||||.....:|||||.....:|||||.....:
244 TTCCTCAGCAGACCTCTGATTCGGGTGGAGAAATTTCTTTTGTAGT 195
370 uphealavalasnleuleuglyleuleuphehyrleullephglycGlna 387
|||||.....:|||||.....:|||||.....:|||||.....:
194 GTTTCGAGTTAACCTACTGCTTAATCATCTACCTGCTTTGGGAAAG 145
387 laaspvalglnluptrpalalaGlnaArgLysleuthrargleu 401
|||||.....:|||||.....:|||||.....:|||||.....:
144 CAGATCTTCAAGATGGCTGAGAGAGAACTCACTCGTTTG 101

seq_name: gb_est12:AA276173

seq_documentation_block:
LOCUS AA276173 495 bp mRNA EST 01-APR-1997
DEFINITION vc31a07.r1 Barstead MPLRBI Mus musculus cDNA clone IMAGE:776148 5'
similar to TR:G887523 G887523 SODIUM-PHOSPHATE TRANSPORT SYSTEM 1.
; mRNA sequence.
ACCESSION AA276173
VERSION AA276173.1 GI:1918811
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 495)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The Washu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1292357.
Contact: Marra M/Mouse EST Project
Washu-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
MG1:469004
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 459.
Location/Qualifiers
1..495
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="IMAGE:776148"
/clone_11b="Barstead MPLRBI"
/sex="mixed"
/tissue_type="Kidney"
/dev_stage="6 weeks"
/lab_host="DH10B"
/note="Vector: p77T3D-Pac (Pharmacia) with a modified
polylinker. Site 1: EcoRI; Site 2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCGATCGAATCGAAGTGGAGCGGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[CATGATTCGTAAC], digested with Not I and cloned into the
Not I and Eco RI sites of the modified p77T3 vector.
Library constructed by Bob Barstead."

BASE COUNT 124 a 129 c 88 g 154 t
ORIGIN
```

```
alignment_scores:
Quality: 575.00 Length: 164
Ratio: 3.859 Gaps: 0
Percent Similarity: 90.854 Percent Identity: 67.073

alignment_block:
US-09-391-958-1 x AA276173 ..

Align seg 1/1 to: AA276173 from: 1 to: 495

157 AspAspProValSerTyrProThrPheSerThrsGlnuSgluTyrIle 173
|||||.....:|||||.....:|||||.....:|||||.....:
3 GATGACCTGCTCTCTACCCATGATGATGAGCCCAAGAGAAATATAT 52
53 TTTATCCTCTCTGACACACAGTTCACTCAGACAGAACAGACACTTCCA 102
173 eileSerleuLysGlnGlnValGlySerSerlyGlnProleuProI 190
|||||.....:|||||.....:|||||.....:|||||.....:
53 TTTATCCTCTCTGACACACAGTTCACTCAGACAGAACAGACACTTCCA 102
190 leIysAlaMetleuArgSerleuProIleTrrPserIleCysleuGlyCys 206
|||||.....:|||||.....:|||||.....:|||||.....:
103 TCAAGCTATGCTCAATCTCTGCTCTGTCGTCATGTCCTCTGCAC 152
207 PheSerHisGlnTrpLeuValSerThrMetValValTyrIleProThrTy 223
|||||.....:|||||.....:|||||.....:|||||.....:
153 ATGACCCATCATGCTGCTTGAACACCTTATATATGATCACTCCACTA 202
223 rIleSerSerValTyrHisValAsnIleArgAspAsnGlyLeuLeuSera 240
|||||.....:|||||.....:|||||.....:|||||.....:
203 CACACAGTTCTGTATTCAAAGTTAATCATCAGACAAATGGGTCTGCTT 252
240 laleupropheillevalaIatrrpalleleGlyMetValGlyGlyTyrleu 256
|||||.....:|||||.....:|||||.....:|||||.....:
253 CTCTCCCTTATATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
257 AlaAspPheleuThrLysLysPheArgleuIleThrValArgLysIle 273
|||||.....:|||||.....:|||||.....:|||||.....:
303 GCAGATTTCTACTGACAGCAAGAAATTTAGGCTCTTAACCTGAGAAATTT 352
273 ealatrIleleuGlySerleuProSerSerAlaIleuIleValSerleup 290
|||||.....:|||||.....:|||||.....:|||||.....:
353 CACACAGCTTTTAGGAATGCTCTCTCTGACAGCCCTGTTGGGCTCTGC 401
290 roTyrleuAsnSerGlyTyrIleThrAlaThrAlaIleuThrLeuSer 306
|||||.....:|||||.....:|||||.....:|||||.....:
402 C.TATATCCCAATCCAGCTATATTTACAACAATTATATTCTGCACAAATTTCC 451
307 CysGlyLeuSerThrIleuCysGlnSerGlyTyrIleThrIleAsn 320
|||||.....:|||||.....:|||||.....:|||||.....:
452 TGTGACTGTGCCCTCTATCTCAGGACGAGAACTATATTTAT 493

seq_name: gb_est27:AI427938

seq_documentation_block:
LOCUS AI427938 452 bp mRNA EST 09-MAR-1999
DEFINITION mm25d12.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:52251 3' similar to TR:000476 000476 SODIUM PHOSPHATE
TRANSPORTER. ; mRNA sequence.
ACCESSION AI427938
VERSION AI427938.1 GI:4273864
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 452)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE The Washu-HMI Mouse EST Project 1999
```

JOURNAL Unpublished (1999)
On Feb 17, 1998 this sequence version replaced g1:2889559.
COMMENT Contact: Maria M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 400.

FEATURES

source
1..452
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:52251"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1:
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

BASE COUNT 113 a 119 c 79 g 140 t 1 others

ORIGIN

alignment_scores: Quality: 546.00 Length: 149
Ratio: 3.985 Gaps: 0

Percent Similarity: 91.946 Percent Identity: 66.443

alignment_block:
US-09-391-958-1 x A1427938 ..

Align seg 1/1 to: A1427938 from: 1 to: 452

```

158 ASPROVALSERYRPTPILESERTherSerLulysGluYrIleI 174
|||||
6 GACCTGTCTCAACCATGATAGTGGCCCAAGAAAGATATATTTT 55
|||||
174 eserSerLeuLysGlnGlnValGlySerSerLysGlnProLeuProIleT 191
|||||
56 ATCCCTCTGAACCAACAGTTCAGCTCAGAGACAGCACTTCCATCA 105
|||||
191 ysAlaMetLeuArgSerLeuProIleTTPSerLecysLeuGlyCysPhe 207
|||||
106 AAGTATATCTCAAAATCTGCTCTGCTGCTCAGTGTCTGCACTATG 155
|||||
208 SerHISGlnTPLeuValSerTherMetValYrIleProThYrI 224
|||||
156 ACCCAATGAGTGGCTGTAAACCTTTATATGTAACACTCCACACTCAT 205
|||||
224 eserSerValYrHISValAsnIleArgAspAsnGlyLeuLeuSerAlaL 241
|||||
206 CAGTCTGTATTCAAAGTTAACATCAGACACATGGTTCCTGCTTCTC 255
|||||
241 euPProheLeuValAlaTPValIleGlyMetValGlyGlyTyrLeuAla 257
|||||
256 TTCCCTTTATTTGTGCTGGGCTTGTGATCTGGGAGGCTGGCTGGCA 305
|||||
258 AspPheLeuLeuThrLysLysPheArgLeuLeuIleThrValArgLysIleAl 274
|||||
306 GATTTTCTACTGACGAAGATTTTAGGCTCATTAACACTGTGAGAAATTCAT 355

```

```

274 aThrIleLeuGlySerLeuProSerSerAlaLeuIleValSerLeuProT 291
|||||
356 CACACTTTTAGGAATGCTCTCTCTGACGCCCTGTTGGCGCTCTGCCCT 405
|||||
291 YrLeuAsnSerGlyYrIleThrAlaThrAlaLeuLeuThrLeuSer 306
|||||
406 ATATCCATCCAGCATATATACACATATATATTTCTGACAAATTCN 452
|||||

```

seq_name: gb.est3:AA073942

seq_documentation_block: 455 bp

LOCUS AA073942 mRNA

DEFINITION mm97h11.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone

IMAGE:536421 5' similar to TR:G165690 G165690 RENCAL CORTICAL

NA/P-I-GOTRANSFORMER.; mRNA sequence.

AA073942

VERSION AA073942.1 GI:1595689

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Euetheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 455)

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,

Wetstein,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The Washu-HMI Mouse EST Project

Unpublished (1996)

On Sep 12, 1996 this sequence version replaced g1:1400896.

Contact: Maria M/Mouse EST Project

Washu-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LML; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:323357

Seg primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 447.

FEATURES

source

1..453

/organism="Mus musculus"

/strain="NIH/Swiss"

/db_xref="taxon:10090"

/clone="IMAGE:536421"

/clone_lib="Stratagene mouse heart (#937316)"

/sex="pooled"

/tissue_type="heart"

/dev_stage="13 day embryos"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: heart; Vector: pBluescript SK-; Site: 1:

EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.

Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'

adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

alignment_scores:	511.00	Length: 160
Quality:		Gaps: 1
Ratio:	3.961	
Percent Similarity:	80.625	Percent Identity: 60.625
alignment_block:		

Clone distribution: NCI-CGAP clone distribution

109 a	134 c	92 g	144 c	2000
-------	-------	------	-------	------

Wyllie, T., Lennon, G., Pearce, J., Theising, B., Waterston, R.

109 a	134 c	92 g	144 c	2000
-------	-------	------	-------	------

TITLE The Washu-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394568.
Contact: Marra M/Mouse EST Project

Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LINDA; contact the
IMAGE Consortium (info@image.lind.gov) for further information.
MGI:316399

Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 314.
Location/Qualifiers

FEATURES
SOURCE 1..438
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_image="52251"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

BASE COUNT 140 a 94 c 98 g 106 t
ORIGIN

alignment_scores:

Quality: 424.00 Length: 113
Ratio: 4.198 Gaps: 0
Percent Similarity: 89.381 Percent Identity: 69.912

alignment_block:
US-09-391-958-1 x AA068000/rev ..

Align seg 1/1 to reverse of: AA068000 from: 1 to: 438

```

289 leuprotyleuansserglytyrilethralathralaleuThrle 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 CTCGCCATATCCATCCACCTATATACACATTTATCTCTGACAAAT 389
305 usercysglyleuserthreucysglnserglyiletyrileasval 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 TTCCTGTGAGCTGTGCCCTCATCTACGACGACGAACTATATTAATCGT 339
322 euaspllealaproargtyrsersepherleuemetglyalaserargly 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 TAGACATTTGCTCCAGGATGCGAGCTTCTCATGGGAAACATCAAGAGA 289
339 pheserterlealaprovalillevaliprorthrvalserglyphelau 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
288 TTGGCACTTCATCGCGCTGTGCTGTACCATTTGCTGCTTTTCTTCT 239
355 userglaspproglnphelglytrpargaanvalphaperleuLeuDea 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 CACCCAGAGACTCTGAATTCGCTGAGGAAATTTCTTTTCTAGCTTTG 189
372 lavalasleuleuglyleuLeuDeaTherleuilepheglyglualasp 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 CAGTTAACCTACTGCTTATCATCTACCTCGCTTGGGAAAGACAGAT 139
389 ValGlnGluTrpAlaLysGluArgLysLeuTharGlu 401
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

138 GTTCAGAAATGGCTAGAGAGAGAGAACTCATCTGTTG 100
seq_name: gp_est18:AA675103

seq_documentation_block:
LOCUS AA675103 462 bp mRNA EST 28-NOV-1997
DEFINITION vq03e06.r1 stratagene mouse diaphragm (#937303) Mus musculus cDNA
clone IMAGE:1093186 5' similar to TR:G450532 G450532
SODIUM-PHOSPHATE TRANSPORT SYSTEM 1. ; mRNA sequence.

ACCESSION AA675103 GI:2652340
VERSION AA675103.1 GI:2652340
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The Washu-HHMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1405024.
Contact: Marra M/Mouse EST Project
Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LINDA; contact the
IMAGE Consortium (info@image.lind.gov) for further information.
MGI:599418
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 433.
Location/Qualifiers

FEATURES
SOURCE 1..462
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="1093186"
/clone_lib="Stratagene mouse diaphragm (#937303)"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dT. Average
insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3' "

BASE COUNT 102 a 120 c 108 g 131 t 1 others
ORIGIN

alignment_scores:
Quality: 416.00 Length: 179
Ratio: 3.382 Gaps: 2
Percent Similarity: 68.715 Percent Identity: 48.045

alignment_block:
US-09-391-958-1 x AA675103 ..

Align seg 1/1 to: AA675103 from: 1 to: 462

```

96 TrpGlyProProlGlnGluArgSerArgLeuCysSerIleAlaLeuSerG 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 TTGGCTCCGCACTTGAGACGACGACGCTCACACATGCGTGCTCA.. 49
112 ymetleuLeuGlyCysPheThrAlaIleLeuIleGlyGlyPheIleSer 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49 ..... 49

```

```

129 IuThleuCluYrProphValPheTYrIlePheGlyGlyValGlyCys 145
    |||:|||||
50 .....GTAAGCATTTGGCTGT 64
146 ValCysCysLeuLeuTrpPheValIleTYrAspAspProValSerTY 162
    |||:|||||
65 GTCTGCTGTCCTGTGGTTCACGGTGAATTATGATGACCCCATGATCA 114
162 rProTPPIleSerThrSerGluLysGluTYrIleLeuSerSerLeuLysG 179
    :||| |||:|||||
115 CCCATGCAATAGTGTGAGGAAAAGAACACATCATCTTCAGTGGCTC 164
179 InGlnValGlySerSerLysGlnProLeuProIleLysAlaMetLeuArg 195
    |||:|||||
165 AGCAGTGCAGTCTCCACAGAGGTCTGTCCCATTAAGGCGATGTCAGA 214
196 SerLeuProIleTrpSerIleCysLeuGlyCysPheSerHisGlnTrpLe 212
    |||:|||||
215 TGCTGGCCACTGTGGCCATTTCATGGGTTTTCAGTCATTTCTGGCT 264
212 uValSerThrMetValValTYrIleProThrTYrIleSerSerValTYrH 229
    | :|:|||||
265 TTG.ACCATATCATTAACATACCTACGACGATCATCAGCACAGTGTGC 313
229 IsValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuProPheIleVal 245
    |||:|||||
314 ACGTTAATCATCAGAGACAGTGGGTCTGTGCTCTCCCTTCATTGCT 363
246 AlaTrpValIleGlyMetValGlyTYrLeuAlaAspPheLeuLeuTh 262
    |||:|||||
364 GCCTCAAGCTGACAAATTTAGAGAGTCAGATGGCAGATTTCCTTCCTC 413
262 rLysLys...PheArgLeuIleThrValArgLysIle 273
    :|:|||||
414 CAGGAATCTTTCAGCTTAATCATCAGTTGANAACTC 450

```